PROVISIONAL APPLICATION COVER SHEET

This is a request for filing a PROVISIONAL APPLICATION under 37 CFR 1.53(b)(2)

"EXPRESS MAIL" MAILING LABEL
Number: EM 334 768 213 US Date of Deposit: October 29,1997
I hereby certify that this paper or fee is being deposited with the United States Postal Service "Express Mail Post Office to Addressee" service under 37 C.F.R. 1.10 on the date indicated above and is addressed to the Assistant Commissioner for Patents, Washington, D.C. 20231.
Typed or Printed Name: Teri Rynshaw
Signed:

Docket Number:	P-65409/WHD/MTK (Genentech Docket No. PO1188)	Type a plus sign (+) inside this box →	+

INVENTOR(s)/APPLICANT(s)

LAST NAME	FIRST NAME	MIDDLE INITIAL	RESIDENCE (CITY AND EITHER STATE OR FOREIGN COUNTRY)
WOOD	WILLIAM	I.	San Mateo, California
GURNEY	AUSTIN	L.	Belmont, California
GODDARD	AUDREY		San Francisco, California

TITLE OF THE INVENTION (*280 characters max*)

NOVEL POLYPEPTIDES HAVING HOMOLOGY TO INTEGRIN AND NUCLEIC ACIDS ENCODING THEREFOR

CORRESPONDENCE ADDRESS

Walter H. Dreger, Registration No. 24, 190
FLEHR HOHBACH TEST ALBRITTON & HERBERT LLP, 4 Embarcadero Center, Suite 3400, San Francisco

STATE	CA	ZIP CODE	94111	COUNTRY	US

ENCLOSED APPLICATION PARTS (*check all that apply*)

X	Specification	*Number of Pages*	39		Small Entity Statement
X	Drawings	*Number of Sheets*	35		Other (*specify*):

METHOD OF PAYMENT OF FILING FEES FOR THIS PROVISIONAL APPLICATION FOR PATENT (*check one*)

X	A check or money order is enclosed to cover the Provisional filing fees		PROVISIONAL FILING FEE AMOUNT ($)	$150
X	The Commissioner is hereby authorized to charge any underpayment or overpayment to Deposit Account Number:	06-1300 (Order No. P-65409/WHD/MTK)		

The invention was made by an agency of the United States Government or under a contract with an agency of the United States Government.

X No.

☐ Yes, the name of the U.S. Government Agency and the Government contract number are: _____

Respectfully submitted,

SIGNATURE: By: Jan P. Brunelle Reg No. 35,081 Date Oct. 29, 1997

For: TYPED or PRINTED NAME Walter H. Dreger REGISTRATION NO. (*if appropriate*) 24,190

☐ Additional inventors are being named on separately numbered sheets attached hereto.

USE ONLY FOR FILING A PROVISIONAL APPLICATION FOR PATENT

Burden Hour Statement: This form is estimated to take .2 hours to complete. Times will vary depending upon the needs of the individual case. Any comments on the amount of time you are required to complete this form should be sent to the Chief Information Officer, Assistant Commissioner for Patents, Washington, DC 20231. DO NOT SEND FEES OR COMPLETED FORMS TO THIS ADDRESS. SEND TO: Box Provisional Application, Assistant Commissioner for Patents, Washington, D.C. 20231.

PATENT
Attorney Docket No.: P-65409/WHD/LRJ
(Genentech Docket No.: 1188)

NOVEL POLYPEPTIDES HAVING HOMOLOGY TO INTEGRIN AND NUCLEIC ACIDS ENCODING THEREFOR

FIELD OF THE INVENTION

The present invention relates generally to the identification and isolation of novel DNA and to the recombinant production of novel polypeptides having homology to integrin, designated herein as "PRO295" polypeptides.

BACKGROUND OF THE INVENTION

Extracellular proteins play an important role in the formation, differentiation and maintenance of multicellular organisms. The fate of many individual cells, e.g., proliferation, migration, differentiation, or interaction with other cells, is typically governed by information received from other cells and/or the immediate environment. This information is often transmitted by secreted polypeptides (for instance, mitogenic factors, survival factors, cytotoxic factors, differentiation factors, neuropeptides, and hormones) which are, in turn, received and interpreted by diverse cell receptors or membrane-bound proteins. These secreted polypeptides or signaling molecules normally pass through the cellular secretory pathway to reach their site of action in the extracellular environment.

Secreted proteins have various industrial applications, including pharmaceuticals, diagnostics, biosensors and bioreactors. Most protein drugs available at present, such as thrombolytic agents, interferons, interleukins, erythropoietins, colony stimulating factors, and various other cytokines, are secretory proteins. Their receptors, which are membrane proteins, also have potential as therapeutic or diagnostic agents. Efforts are being undertaken by both industry and academia to identify new, native secreted proteins. Many efforts are focused on the screening of mammalian recombinant DNA libraries to identify the coding sequences for novel secreted proteins. Examples of screening methods and techniques are described in the literature [see, for example, Klein et al., Proc. Natl. Acad. Sci., 93:7108-7113 (1996); U.S.

Patent No. 5,536,637)].

The integrins comprise a supergene family of cell-surface glycoprotein receptors that promote cellular adhesion. Each cell has numerous receptors that define its cell adhesive capabilities. Integrins are involved in a wide variety of interaction between cells and other cells or matrix components. The integrins are of particular importance in regulating movement and function of immune system cells The platelet IIb/IIIA integrin complex is of particular importance in regulating platelet aggregation. A member of the integrin family, integrin β-6, is expressed on epithelial cells and modulates epithelial inflammation. Another integrin, leucocyte-associated antigen-1 (LFA-1) is important in the adhesion of lymphocytes during an immune response. The integrins are expressed as heterodimers of non-covalently associated alpha and beta subunits. Given the physiological importance of cell adhesion mechanisms *in vivo*, efforts are currently being under taken to identify new, native proteins which are involved in cell adhesion. We describe herein the identification of a novel polypeptide which has homology to integrin.

SUMMARY OF THE INVENTION

Applicants have identified a cDNA clone that encodes a novel polypeptide, wherein the polypeptide is designated in the present application as "PRO295".

In one embodiment, the invention provides an isolated nucleic acid molecule comprising DNA encoding a PRO295 polypeptide. In one aspect, the isolated nucleic acid comprises DNA encoding the PRO295 polypeptide having amino acid residues 1 to 350 of Figure 2 (SEQ ID NO:3), or is complementary to such encoding nucleic acid sequence, and remains stably bound to it under at least moderate, and optionally, under high stringency conditions.

In another embodiment, the invention provides a vector comprising DNA encoding a PRO295 polypeptide. A host cell comprising such a vector is also provided. By way of example, the host cells may be CHO cells, *E. coli*, or yeast. A process for producing PRO295 polypeptides is further provided and comprises culturing host cells under conditions suitable for expression of PRO295 and recovering PRO295 from the cell culture.

In another embodiment, the invention provides isolated PRO295 polypeptide. In particular, the invention provides isolated native sequence PRO295 polypeptide, which in one

embodiment, includes an amino acid sequence comprising residues 1 to 350 of Figure 2 (SEQ ID NO:3).

In another embodiment, the invention provides chimeric molecules comprising a PRO295 polypeptide fused to a heterologous polypeptide or amino acid sequence. An example of such a chimeric molecule comprises a PRO295 polypeptide fused to an epitope tag sequence or a Fc region of an immunoglobulin.

In another embodiment, the invention provides an antibody which specifically binds to a PRO295 polypeptide. Optionally, the antibody is a monoclonal antibody.

BRIEF DESCRIPTION OF THE DRAWINGS

Figure 1 shows a nucleotide sequence (SEQ ID NO:1) containing the nucleotide sequence (SEQ ID NO:2) of a native sequence PRO295 cDNA (nucleotides 153-1202), wherein the nucleotide sequence (SEQ ID NO:1) is designated herein as "UNQ258" and/or "DNA38268". Also presented is the position of the initiator methionine residue. The putative transmembrane domain of the protein is encoded by nucleotides beginning at nucleotide 1586 in the figure.

Figure 2 shows the amino acid sequence (SEQ ID NO:3) derived from nucleotides 153-1202 of the nucleotide sequence shown in Figure 1.

Figures 3A-3B show an alignment of portions of the nucleotide sequences from a variety of expressed sequence tags as well as an intermediate consensus nucleotide sequence assembled therefrom designated as "DNA35814".

Figures 4A through 4AB show an alignment of portions of the nucleotide sequences from a variety of expressed sequence tags as well as an extended consensus nucleotide sequence assembled therefrom using repeated cycles of BLAST and phrap.

Figure 5 shows a BLAST sequence alignment analysis of portions of the PRO295 amino acid sequence derived from the DNA38268 molecule ("DNA38268") with portions of the integrin beta-6 subunit - cavia porcellus ("ITB6_CAVPO").

Figure 6 shows a BLAST sequence alignment analysis of portions of the PRO295 amino acid sequence derived from the DNA38268 molecule ("DNA38268") with portions of the lens fiber protein clfest4 precursor - gallus ("LFE4_CHICK").

DETAILED DESCRIPTION OF THE PREFERRED EMBODIMENTS

I. Definitions

The terms "PRO295 polypeptide" and "PRO295" when used herein encompass native sequence PRO295 and PRO295 polypeptide variants (which are further defined herein). The PRO295 polypeptides may be isolated from a variety of sources, such as from human tissue types or from another source, or prepared by recombinant or synthetic methods.

A "native sequence PRO295 polypeptide" comprises a polypeptide having the same amino acid sequence as a PRO295 polypeptide derived from nature. Such native sequence PRO295 polypeptide can be isolated from nature or can be produced by recombinant or synthetic means. The term "native sequence PRO295 polypeptide" specifically encompasses naturally-occurring truncated or secreted forms of a PRO295 polypeptide (*e.g.*, an extracelluldr domain sequence), naturally-occurring variant forms (*e.g.*, alternatively spliced forms) and naturally-occurring allelic variants of a PRO295 polypeptide. In one embodiment of the invention, the native sequence PRO295 polypeptide is a mature or full-length native sequence PRO295 polypeptide comprising amino acids 1 to 350 of Figure 2 (SEQ ID NO:3).

"Percent (%) amino acid sequence identity" with respect to the PRO295 amino acid sequences identified herein is defined as the percentage of amino acid residues in a candidate sequence that are identical with the amino acid residues in a PRO295 polypeptide sequence, after aligning the sequences and introducing gaps, if necessary, to achieve the maximum percent sequence identity, and not considering any conservative substitutions as part of the sequence identity. Alignment for purposes of determining percent amino acid sequence identity can be achieved in various ways that are within the skill in the art, for instance, using publicly available computer software such as ALIGN or Megalign (DNASTAR) software. Those skilled in the art can determine appropriate parameters for measuring alignment, including any algorithms needed to achieve maximal alignment over the full length of the sequences being compared.

"Percent (%) nucleic acid sequence identity" with respect to the UNQ258 (DNA38268) sequence (SEQ ID NO:1) and nucleotides 153-1202 of the sequnece shown in Figure 1 (SEQ ID NO:2) is defined as the percentage of nucleotides in a candidate sequence that are identical with the nucleotides in the UNQ258 (DNA38268) sequence (SEQ ID NO:1) or nucleotides 153-1202 of the sequence shown in Figure 1 (SEQ ID NO:2), respectively, after aligning the

sequences and introducing gaps, if necessary, to achieve the maximum percent sequence identity. Alignment for purposes of determining percent nucleic acid sequence identity can be achieved in various ways that are within the skill in the art, for instance, using publicly available computer software such as ALIGN or Megalign (DNASTAR) software. Those skilled in the art can determine appropriate parameters for measuring alignment, including any algorithms needed to achieve maximal alignment over the full length of the sequences being compared.

"Isolated," when used to describe the various polypeptides disclosed herein, means polypeptide that has been identified and separated and/or recovered from a component of its natural environment. Contaminant components of its natural environment are materials that would typically interfere with diagnostic or therapeutic uses for the polypeptide, and may include enzymes, hormones, and other proteinaceous or non-proteinaceous solutes. In preferred embodiments, the polypeptide will be purified (1) to a degree sufficient to obtain at least 15 residues of N-terminal or internal amino acid sequence by use of a spinning cup sequenator, or (2) to homogeneity by SDS-PAGE under non-reducing or reducing conditions using Coomassie blue or, preferably, silver stain. Isolated polypeptide includes polypeptide *in situ* within recombinant cells, since at least one component of the PRO295 polypeptide natural environment will not be present. Ordinarily, however, isolated polypeptide will be prepared by at least one purification step.

An "isolated" PRO295 polypeptide-encoding nucleic acid molecule is a nucleic acid molecule that is identified and separated from at least one contaminant nucleic acid molecule with which it is ordinarily associated in the natural source of the PRO295 polypeptide-encoding nucleic acid. An isolated PRO295 polypeptide-encoding nucleic acid molecule is other than in the former setting in which it is found in nature. Isolated PRO295 polypeptide-encoding nucleic acid molecules therefore are distinguished from the PRO295 polypeptide-encoding nucleic acid molecule as it exists in natural cells. However, an isolated PRO295 polypeptide-encoding nucleic acid molecule includes PRO295 polypeptide-encoding nucleic acid molecules contained in cells that ordinarily express PRO295 polypeptide where, for example, the nucleic acid molecule is in a chromosomal location different from that of natural cells.

The term "control sequences" refers to DNA sequences necessary for the expression

of an operably linked coding sequence in a particular host organism. The control sequences that are suitable for prokaryotes, for example, include a promoter, optionally an operator sequence, and a ribosome binding site. Eukaryotic cells are known to utilize promoters, polyadenylation signals, and enhancers.

Nucleic acid is "operably linked" when it is placed into a functional relationship with another nucleic acid sequence. For example, DNA for a presequence or secretory leader is operably linked to DNA for a polypeptide if it is expressed as a preprotein that participates in the secretion of the polypeptide; a promoter or enhancer is operably linked to a coding sequence if it affects the transcription of the sequence; or a ribosome binding site is operably linked to a coding sequence if it is positioned so as to facilitate translation. Generally, "operably linked" means that the DNA sequences being linked are contiguous, and, in the case of a secretory leader, contiguous and in reading phase. However, enhancers do not have to be contiguous. Linking is accomplished by ligation at convenient restriction sites. If such sites do not exist, the synthetic oligonucleotide adaptors or linkers are used in accordance with conventional practice.

The term "antibody" is used in the broadest sense and specifically covers single anti-PRO295 polypeptide monoclonal antibodies (including agonist, antagonist, and neutralizing antibodies) and anti-PRO295 antibody compositions with polyepitopic specificity. The term "monoclonal antibody" as used herein refers to an antibody obtained from a population of substantially homogeneous antibodies, *i.e.*, the individual antibodies comprising the population are identical except for possible naturally-occurring mutations that may be present in minor amounts.

"Active" or "activity" for the purposes herein refers to form(s) of PRO295 which retain the biologic and/or immunologic activities of native or naturally-occurring PRO295 polypeptide.

II. Compositions and Methods of the Invention

A. Full-length PRO295 Polypeptide

The present invention provides newly identified and isolated nucleotide sequences encoding polypeptides referred to in the present application as PRO295. In particular, Applicants have identified and isolated cDNA encoding a PRO295 polypeptide, as disclosed

in further detail in the Examples below. Using BLAST and FastA sequence alignment computer programs, Applicants found that various portions of the PRO295 polypeptide have significant homology with the integrin protein as shown in Figures 5 and 6). Accordingly, it is presently believed that PRO295 polypeptide disclosed in the present application is a newly identified member of the integrin family and possesses cell adhesion typical of the integrin family.

B. Modifications of PRO295

Covalent modifications of PRO295 polypeptides are included within the scope of this invention. One type of covalent modification includes reacting targeted amino acid residues of a PRO295 polypeptide with an organic derivatizing agent that is capable of reacting with selected side chains or the N-or C-terminal residues of a PRO295 polypeptide. Derivatization with bifunctional agents is useful, for instance, for crosslinking PRO295 to a water-insoluble support matrix or surface for use in the method for purifying anti-PRO295 antibodies, and vice-versa. Commonly used crosslinking agents include, *e.g.*, 1,1-bis(diazo-acetyl)-2-phenylethane, glutaraldehyde, N-hydroxysuccinimide esters, for example, esters with 4-azidosalicylic acid, homobifunctional imidoesters, including disuccinimidyl esters such as 3,3'-dithiobis(succinimidylpropionate), bifunctional maleimides such as bis-N-maleimido-1,8-octane and agents such as methyl-3-[(p-azidophenyl)dithio]propioimidate.

Other modifications include deamidation of glutaminyl and asparaginyl residues to the corresponding glutamyl and aspartyl residues, respectively, hydroxylation of proline and lysine, phosphorylation of hydroxyl groups of seryl or threonyl residues, methylation of the α-amino groups of lysine, arginine, and histidine side chains [T.E. Creighton, Proteins: Structure and Molecular Properties, W.H. Freeman & Co., San Francisco, pp. 79-86 (1983)], acetylation of the N-terminal amine, and amidation of any C-terminal carboxyl group.

Another type of covalent modification of the PRO295 polypeptide included within the scope of this invention comprises altering the native glycosylation pattern of the polypeptide. "Altering the native glycosylation pattern" is intended for purposes herein to mean deleting one or more carbohydrate moieties found in native sequence PRO295 polypeptide, and/or adding one or more glycosylation sites that are not present in the native sequence PRO295 polypeptide.

Addition of glycosylation sites to PRO295 polypeptides may be accomplished by altering the amino acid sequence thereof. The alteration may be made, for example, by the addition of, or substitution by, one or more serine or threonine residues to the native sequence PRO295 polypeptide (for O-linked glycosylation sites). The PRO295 amino acid sequence may optionally be altered through changes at the DNA level, particularly by mutating the DNA encoding the PRO295 polypeptide at preselected bases such that codons are generated that will translate into the desired amino acids.

Another means of increasing the number of carbohydrate moieties on the PRO295 polypeptide is by chemical or enzymatic coupling of glycosides to the polypeptide. Such methods are described in the art, e.g., in WO 87/05330 published 11 September 1987, and in Aplin and Wriston, CRC Crit. Rev. Biochem., pp. 259-306 (1981).

Removal of carbohydrate moieties present on the PRO295 polypeptide may be accomplished chemically or enzymatically or by mutational substitution of codons encoding for amino acid residues that serve as targets for glycosylation. Chemical deglycosylation techniques are known in the art and described, for instance, by Hakimuddin, et al., Arch. Biochem. Biophys., 259:52 (1987) and by Edge et al., Anal. Biochem., 118:131 (1981). Enzymatic cleavage of carbohydrate moieties on polypeptides can be achieved by the use of a variety of endo-and exo-glycosidases as described by Thotakura et al., Meth. Enzymol., 138:350 (1987).

Another type of covalent modification of PRO295 comprises linking the PRO295 polypeptide to one of a variety of nonproteinaceous polymers, *e.g.*, polyethylene glycol, polypropylene glycol, or polyoxyalkylenes, in the manner set forth in U.S. Patent Nos. 4,640,835; 4,496,689; 4,301,144; 4,670,417; 4,791,192 or 4,179,337.

PRO295 polypeptides of the present invention may also be modified in a way to form chimeric molecules comprising a PRO295 polypeptide fused to another, heterologous polypeptide or amino acid sequence. In one embodiment, such a chimeric molecule comprises a fusion of a PRO295 polypeptide with a tag polypeptide which provides an epitope to which an anti-tag antibody can selectively bind. The epitope tag is generally placed at the amino-or carboxyl-terminus of the PRO295 polypeptide. The presence of such epitope-tagged forms of a PRO295 polypeptide can be detected using an antibody against the tag polypeptide. Also, provision of the epitope tag enables the PRO295 polypeptide to be readily purified by affinity

purification using an anti-tag antibody or another type of affinity matrix that binds to the epitope tag. In an alternative embodiment, the chimeric molecule may comprise a fusion of a PRO295 polypeptide with an immunoglobulin or a particular region of an immunoglobulin. For a bivalent form of the chimeric molecule, such a fusion could be to the Fc region of an IgG molecule.

Various tag polypeptides and their respective antibodies are well known in the art. Examples include poly-histidine (poly-his) or poly-histidine-glycine (poly-his-gly) tags; the flu HA tag polypeptide and its antibody 12CA5 [Field et al., Mol. Cell. Biol., 8:2159-2165 (1988)]; the c-myc tag and the 8F9, 3C7, 6E10, G4, B7 and 9E10 antibodies thereto [Evan et al., Molecular and Cellular Biology, 5:3610-3616 (1985)]; and the Herpes Simplex virus glycoprotein D (gD) tag and its antibody [Paborsky et al., Protein Engineering, 3(6):547-553 (1990)]. Other tag polypeptides include the Flag-peptide [Hopp et al., BioTechnology, 6:1204-1210 (1988)]; the KT3 epitope peptide [Martin et al., Science, 255:192-194 (1992)]; an α-tubulin epitope peptide [Skinner et al., J. Biol. Chem., 266:15163-15166 (1991)]; and the T7 gene 10 protein peptide tag [Lutz-Freyermuth et al., Proc. Natl. Acad. Sci. USA, 87:6393-6397 (1990)].

C. Preparation of PRO295

The description below relates primarily to production of PRO295 by culturing cells transformed or transfected with a vector containing at least nucleotides 153-1202 (SEQ ID NO:2) of the UNQ258 (DNA38268) nucleic acid (SEQ ID NO:1). It is, of course, contemplated that alternative methods, which are well known in the art, may be employed to prepare PRO295 polypeptides. For instance, the PRO295 sequence, or portions thereof, may be produced by direct peptide synthesis using solid-phase techniques [see, e.g., Stewart et al., Solid-Phase Peptide Synthesis, W.H. Freeman Co., San Francisco, CA (1969); Merrifield, J. Am. Chem. Soc., 85:2149-2154 (1963)]. *In vitro* protein synthesis may be performed using manual techniques or by automation. Automated synthesis may be accomplished, for instance, using an Applied Biosystems Peptide Synthesizer (Foster City, CA) using manufacturer's instructions. Various portions of PRO295 polypeptides may be chemically synthesized separately and combined using chemical or enzymatic methods to produce a full-length PRO295 polypeptide.

1. Isolation of DNA Encoding PRO295

DNA encoding a PRO295 polypeptide may be obtained from a cDNA library prepared from tissue believed to possess the PRO295 mRNA and to express it at a detectable level. Accordingly, human PRO295-encoding DNA can be conveniently obtained from a cDNA library prepared from human tissue, such as described in the Examples. The PRO295-encoding gene may also be obtained from a genomic library or by oligonucleotide synthesis.

Libraries can be screened with probes (such as antibodies to a PRO295 polypeptide or oligonucleotides of at least about 20-80 bases) designed to identify the gene of interest or the protein encoded by it. Screening the cDNA or genomic library with the selected probe may be conducted using standard procedures, such as described in Sambrook et al., Molecular Cloning: A Laboratory Manual (New York: Cold Spring Harbor Laboratory Press, 1989). An alternative means to isolate the gene encoding PRO295 is to use PCR methodology [Sambrook et al., supra; Dieffenbach et al., PCR Primer:A Laboratory Manual (Cold Spring Harbor Laboratory Press, 1995)].

The Examples below describe techniques for screening a cDNA library. The oligonucleotide sequences selected as probes should be of sufficient length and sufficiently unambiguous that false positives are minimized. The oligonucleotide is preferably labeled such that it can be detected upon hybridization to DNA in the library being screened. Methods of labeling are well known in the art, and include the use of radiolabels like ^{32}P-labeled ATP, biotinylation or enzyme labeling. Hybridization conditions, including moderate stringency and high stringency, are provided in Sambrook et al., supra.

Sequences identified in such library screening methods can be compared and aligned to other known sequences deposited and available in public databases such as GenBank or other private sequence databases. Sequence identity (at either the amino acid or nucleotide level) within defined regions of the molecule or across the full-length sequence can be determined through sequence alignment using computer software programs such as ALIGN, DNAstar, and INHERIT which employ various algorithms to measure homology.

Nucleic acid having protein coding sequence may be obtained by screening selected cDNA or genomic libraries using the deduced amino acid sequence disclosed herein for the first time, and, if necessary, using conventional primer extension procedures as described in

Sambrook et al., supra, to detect precursors and processing intermediates of mRNA that may not have been reverse-transcribed into cDNA.

2. Selection and Transformation of Host Cells

Host cells are transfected or transformed with expression or cloning vectors described herein for PRO295 polypeptide production and cultured in conventional nutrient media modified as appropriate for inducing promoters, selecting transformants, or amplifying the genes encoding the desired sequences. The culture conditions, such as media, temperature, pH and the like, can be selected by the skilled artisan without undue experimentation. In general, principles, protocols, and practical techniques for maximizing the productivity of cell cultures can be found in Mammalian Cell Biotechnology: a Practical Approach, M. Butler, ed. (IRL Press, 1991) and Sambrook et al., supra.

Methods of transfection are known to the ordinarily skilled artisan, for example, $CaPO_4$ and electroporation. Depending on the host cell used, transformation is performed using standard techniques appropriate to such cells. The calcium treatment employing calcium chloride, as described in Sambrook et al., supra, or electroporation is generally used for prokaryotes or other cells that contain substantial cell-wall barriers. Infection with *Agrobacterium tumefaciens* is used for transformation of certain plant cells, as described by Shaw et al., Gene, 23:315 (1983) and WO 89/05859 published 29 June 1989. For mammalian cells without such cell walls, the calcium phosphate precipitation method of Graham and van der Eb, Virology, 52:456-457 (1978) can be employed. General aspects of mammalian cell host system transformations have been described in U.S. Patent No. 4,399,216. Transformations into yeast are typically carried out according to the method of Van Solingen et al., J. Bact., 130:946 (1977) and Hsiao et al., Proc. Natl. Acad. Sci. (USA), 76:3829 (1979). However, other methods for introducing DNA into cells, such as by nuclear microinjection, electroporation, bacterial protoplast fusion with intact cells, or polycations, *e.g.*, polybrene, polyornithine, may also be used. For various techniques for transforming mammalian cells, see Keown et al., Methods in Enzymology, 185:527-537 (1990) and Mansour et al., Nature, 336:348-352 (1988).

Suitable host cells for cloning or expressing the DNA in the vectors herein include prokaryote, yeast, or higher eukaryote cells. Suitable prokaryotes include but are not limited

to eubacteria, such as Gram-negative or Gram-positive organisms, for example, Enterobacteriaceae such as *E. coli*. Various *E. coli* strains are publicly available, such as *E. coli* K12 strain MM294 (ATCC 31,446); *E. coli* X1776 (ATCC 31,537); *E. coli* strain W3110 (ATCC 27,325) and K5 772 (ATCC 53,635).

In addition to prokaryotes, eukaryotic microbes such as filamentous fungi or yeast are suitable cloning or expression hosts for PRO295-encoding vectors. *Saccharomyces cerevisiae* is a commonly used lower eukaryotic host microorganism.

Suitable host cells for the expression of glycosylated PRO295 are derived from multicellular organisms. Examples of invertebrate cells include insect cells such as Drosophila S2 and Spodoptera Sf9, as well as plant cells. Examples of useful mammalian host cell lines include Chinese hamster ovary (CHO) and COS cells. More specific examples include monkey kidney CV1 line transformed by SV40 (COS-7, ATCC CRL 1651); human embryonic kidney line (293 or 293 cells subcloned for growth in suspension culture, Graham et al., J. Gen Virol., 36:59 (1977)); Chinese hamster ovary cells/-DHFR (CHO, Urlaub and Chasin, Proc. Natl. Acad. Sci. USA, 77:4216 (1980)); mouse sertoli cells (TM4, Mather, Biol. Reprod., 23:243-251 (1980)); human lung cells (W138, ATCC CCL 75); human liver cells (Hep G2, HB 8065); and mouse mammary tumor (MMT 060562, ATCC CCL51). The selection of the appropriate host cell is deemed to be within the skill in the art.

3. Selection and Use of a Replicable Vector

The nucleic acid (*e.g.*, cDNA or genomic DNA) encoding the desired PRO295 polypeptide may be inserted into a replicable vector for cloning (amplification of the DNA) or for expression. Various vectors are publicly available. The vector may, for example, be in the form of a plasmid, cosmid, viral particle, or phage. The appropriate nucleic acid sequence may be inserted into the vector by a variety of procedures. In general, DNA is inserted into an appropriate restriction endonuclease site(s) using techniques known in the art. Vector components generally include, but are not limited to, one or more of a signal sequence, an origin of replication, one or more marker genes, an enhancer element, a promoter, and a transcription termination sequence. Construction of suitable vectors containing one or more of these components employs standard ligation techniques which are known to the skilled artisan.

The desired PRO295 polypeptide may be produced recombinantly not only directly, but also as a fusion polypeptide with a heterologous polypeptide, which may be a signal sequence or other polypeptide having a specific cleavage site at the N-terminus of the mature protein or polypeptide. In general, the signal sequence may be a component of the vector, or it may be a part of the PRO295-encoding DNA that is inserted into the vector. The signal sequence may be a prokaryotic signal sequence selected, for example, from the group of the alkaline phosphatase, penicillinase, lpp, or heat-stable enterotoxin II leaders. For yeast secretion the signal sequence may be, *e.g.*, the yeast invertase leader, alpha factor leader (including *Saccharomyces* and *Kluyveromyces* α-factor leaders, the latter described in U.S. Patent No. 5,010,182), or acid phosphatase leader, the *C. albicans* glucoamylase leader (EP 362,179 published 4 April 1990), or the signal described in WO 90/13646 published 15 November 1990. In mammalian cell expression, mammalian signal sequences may be used to direct secretion of the protein, such as signal sequences from secreted polypeptides of the same or related species, as well as viral secretory leaders.

Both expression and cloning vectors contain a nucleic acid sequence that enables the vector to replicate in one or more selected host cells. Such sequences are well known for a variety of bacteria, yeast, and viruses. The origin of replication from the plasmid pBR322 is suitable for most Gram-negative bacteria, the 2μ plasmid origin is suitable for yeast, and various viral origins (SV40, polyoma, adenovirus, VSV or BPV) are useful for cloning vectors in mammalian cells.

Expression and cloning vectors will typically contain a selection gene, also termed a selectable marker. Typical selection genes encode proteins that (a) confer resistance to antibiotics or other toxins, *e.g.*, ampicillin, neomycin, methotrexate, or tetracycline, (b) complement auxotrophic deficiencies, or (c) supply critical nutrients not available from complex media, *e.g.*, the gene encoding D-alanine racemase for *Bacilli*.

An example of suitable selectable markers for mammalian cells are those that enable the identification of cells competent to take up the PRO295-encoding nucleic acid, such as DHFR or thymidine kinase. An appropriate host cell when wild-type DHFR is employed is the CHO cell line deficient in DHFR activity, prepared and propagated as described by Urlaub et al., Proc. Natl. Acad. Sci. USA, 77:4216 (1980). A suitable selection gene for use in yeast is the *trp*1 gene present in the yeast plasmid YRp7 [Stinchcomb et al., Nature, 282:39 (1979);

Kingsman et al., Gene, 7:141 (1979); Tschemper et al., Gene, 10:157 (1980)]. The *trp*1 gene provides a selection marker for a mutant strain of yeast lacking the ability to grow in tryptophan, for example, ATCC No. 44076 or PEP4-1 [Jones, Genetics, 85:12 (1977)].

Expression and cloning vectors usually contain a promoter operably linked to the PRO295-encoding nucleic acid sequence to direct mRNA synthesis. Promoters recognized by a variety of potential host cells are well known. Promoters suitable for use with prokaryotic hosts include the β-lactamase and lactose promoter systems [Chang et al., Nature, 275:615 (1978); Goeddel et al., Nature, 281:544 (1979)], alkaline phosphatase, a tryptophan (trp) promoter system [Goeddel, Nucleic Acids Res., 8:4057 (1980); EP 36,776], and hybrid promoters such as the tac promoter [deBoer et al., Proc. Natl. Acad. Sci. USA, 80:21-25 (1983)]. Promoters for use in bacterial systems also will contain a Shine-Dalgarno (S.D.) sequence operably linked to the DNA encoding the PRO295 polypeptide.

Examples of suitable promoting sequences for use with yeast hosts include the promoters for 3-phosphoglycerate kinase [Hitzeman et al., J. Biol. Chem., 255:2073 (1980)] or other glycolytic enzymes [Hess et al., J. Adv. Enzyme Reg., 7:149 (1968); Holland, Biochemistry, 17:4900 (1978)], such as enolase, glyceraldehyde-3-phosphate dehydrogenase, hexokinase, pyruvate decarboxylase, phosphofructokinase, glucose-6-phosphate isomerase, 3-phosphoglycerate mutase, pyruvate kinase, triosephosphate isomerase, phosphoglucose isomerase, and glucokinase.

Other yeast promoters, which are inducible promoters having the additional advantage of transcription controlled by growth conditions, are the promoter regions for alcohol dehydrogenase 2, isocytochrome C, acid phosphatase, degradative enzymes associated with nitrogen metabolism, metallothionein, glyceraldehyde-3-phosphate dehydrogenase, and enzymes responsible for maltose and galactose utilization. Suitable vectors and promoters for use in yeast expression are further described in EP 73,657.

PRO295 transcription from vectors in mammalian host cells is controlled, for example, by promoters obtained from the genomes of viruses such as polyoma virus, fowlpox virus (UK 2,211,504 published 5 July 1989), adenovirus (such as Adenovirus 2), bovine papilloma virus, avian sarcoma virus, cytomegalovirus, a retrovirus, hepatitis-B virus and Simian Virus 40 (SV40), from heterologous mammalian promoters, *e.g.*, the actin promoter or an immunoglobulin promoter, and from heat-shock promoters, provided such promoters are

compatible with the host cell systems.

Transcription of a DNA encoding a PRO295 polypeptide by higher eukaryotes may be increased by inserting an enhancer sequence into the vector. Enhancers are cis-acting elements of DNA, usually about from 10 to 300 bp, that act on a promoter to increase its transcription. Many enhancer sequences are now known from mammalian genes (globin, elastase, albumin, α-fetoprotein, and insulin). Typically, however, one will use an enhancer from a eukaryotic cell virus. Examples include the SV40 enhancer on the late side of the replication origin (bp 100-270), the cytomegalovirus early promoter enhancer, the polyoma enhancer on the late side of the replication origin, and adenovirus enhancers. The enhancer may be spliced into the vector at a position 5' or 3' to the PRO295 coding sequence, but is preferably located at a site 5' from the promoter.

Expression vectors used in eukaryotic host cells (yeast, fungi, insect, plant, animal, human, or nucleated cells from other multicellular organisms) will also contain sequences necessary for the termination of transcription and for stabilizing the mRNA. Such sequences are commonly available from the 5' and, occasionally 3', untranslated regions of eukaryotic or viral DNAs or cDNAs. These regions contain nucleotide segments transcribed as polyadenylated fragments in the untranslated portion of the mRNA encoding PRO295.

Still other methods, vectors, and host cells suitable for adaptation to the synthesis of PRO295 polypeptides in recombinant vertebrate cell culture are described in Gething et al., Nature, 293:620-625 (1981); Mantei et al., Nature, 281:40-46 (1979); EP 117,060; and EP 117,058.

4. Detecting Gene Amplification/Expression

Gene amplification and/or expression may be measured in a sample directly, for example, by conventional Southern blotting, Northern blotting to quantitate the transcription of mRNA [Thomas, Proc. Natl. Acad. Sci. USA, 77:5201-5205 (1980)], dot blotting (DNA analysis), or *in situ* hybridization, using an appropriately labeled probe, based on the sequences provided herein. Alternatively, antibodies may be employed that can recognize specific duplexes, including DNA duplexes, RNA duplexes, and DNA-RNA hybrid duplexes or DNA-protein duplexes. The antibodies in turn may be labeled and the assay may be carried out where the duplex is bound to a surface, so that upon the formation of duplex on the surface, the presence of antibody bound to the duplex can be detected.

Gene expression, alternatively, may be measured by immunological methods, such as immunohistochemical staining of cells or tissue sections and assay of cell culture or body fluids, to quantitate directly the expression of gene product. Antibodies useful for immunohistochemical staining and/or assay of sample fluids may be either monoclonal or polyclonal, and may be prepared in any mammal. Conveniently, the antibodies may be prepared against a native sequence PRO295 polypeptide or against a synthetic peptide based on the DNA sequences provided herein or against exogenous sequence fused to PRO295-encoding DNA and encoding a specific antibody epitope.

5. Purification of Polypeptide

Forms of PRO295 may be recovered from culture medium or from host cell lysates. If membrane-bound, it can be released from the membrane using a suitable detergent solution (*e.g.* Triton-X 100) or by enzymatic cleavage. Cells employed in expression of PRO295 polypeptides can be disrupted by various physical or chemical means, such as freeze-thaw cycling, sonication, mechanical disruption, or cell lysing agents.

It may be desired to purify PRO295 from recombinant cell proteins or polypeptides. The following procedures are exemplary of suitable purification procedures: by fractionation on an ion-exchange column; ethanol precipitation; reverse phase HPLC; chromatography on silica or on a cation-exchange resin such as DEAE; chromatofocusing; SDS-PAGE; ammonium sulfate precipitation; gel filtration using, for example, Sephadex G-75; protein A Sepharose columns to remove contaminants such as IgG; and metal chelating columns to bind epitope-tagged forms of the PRO295 polypeptide. Various methods of protein purification may be employed and such methods are known in the art and described for example in Deutscher, Methods in Enzymology, 182 (1990); Scopes, Protein Purification:Principles and Practice, Springer-Verlag, New York (1982). The purification step(s) selected will depend, for example, on the nature of the production process used and the particular PRO295 polypeptide produced.

D. Uses for PRO295

Nucleotide sequences (or their complement) encoding PRO295 polypeptides have various applications in the art of molecular biology, including uses as hybridization probes,

in chromosome and gene mapping and in the generation of anti-sense RNA and DNA. PRO295-encoding nucleic acid will also be useful for the preparation of PRO295 polypeptides by the recombinant techniques described herein.

The full-length nucleotide sequence SEQ ID NO:1 or the full-length native sequence PRO295 (SEQ ID NO:2) nucleotide sequence, or portions thereof, may be used as hybridization probes for a cDNA library to isolate the full-length PRO295 gene or to isolate still other genes (for instance, those encoding naturally-occurring variants of PRO295 or PRO295 from other species) which have a desired sequence identity to the PRO295 sequence disclosed in Figure 1 (SEQ ID NO:1). Optionally, the length of the probes will be about 20 to about 50 bases. The hybridization probes may be derived from the UNQ258 (DNA38268) nucleotide sequence of SEQ ID NO:1 as shown in Figure 1 or from genomic sequences including promoters, enhancer elements and introns of native sequence PRO295-encoding DNA. By way of example, a screening method will comprise isolating the coding region of the PRO295 gene using the known DNA sequence to synthesize a selected probe of about 40 bases. Hybridization probes may be labeled by a variety of labels, including radionucleotides such as ^{32}P or ^{35}S, or enzymatic labels such as alkaline phosphatase coupled to the probe via avidin/biotin coupling systems. Labeled probes having a sequence complementary to that of the PRO295 gene of the present invention can be used to screen libraries of human cDNA, genomic DNA or mRNA to determine which members of such libraries the probe hybridizes to. Hybridization techniques are described in further detail in the Examples below.

The probes may also be employed in PCR techniques to generate a pool of sequences for identification of closely related PRO295 sequences.

Nucleotide sequences encoding a PRO295 polypeptide can also be used to construct hybridization probes for mapping the gene which encodes that PRO295 polypeptide and for the genetic analysis of individuals with genetic disorders. The nucleotide sequences provided herein may be mapped to a chromosome and specific regions of a chromosome using known techniques, such as *in situ* hybridization, linkage analysis against known chromosomal markers, and hybridization screening with libraries.

When the coding sequences for PRO295 encode a protein which binds to another protein (example, where the PRO295 polypeptide functions as a receptor), the PRO295 polypeptide can be used in assays to identify the other proteins or molecules involved in the

binding interaction. By such methods, inhibitors of the receptor/ligand binding interaction can be identified. Proteins involved in such binding interactions can also be used to screen for peptide or small molecule inhibitors or agonists of the binding interaction. Also, the receptor PRO295 polypeptide can be used to isolate correlative ligand(s). Screening assays can be designed to find lead compounds that mimic the biological activity of a native PRO295 or a receptor for PRO295. Such screening assays will include assays amenable to high-throughput screening of chemical libraries, making them particularly suitable for identifying small molecule drug candidates. Small molecules contemplated include synthetic organic or inorganic compounds. The assays can be performed in a variety of formats, including protein-protein binding assays, biochemical screening assays, immunoassays and cell based assays, which are well characterized in the art.

Nucleic acids which encode PRO295 polypeptide or any of its modified forms can also be used to generate either transgenic animals or "knock out" animals which, in turn, are useful in the development and screening of therapeutically useful reagents. A transgenic animal (e.g., a mouse or rat) is an animal having cells that contain a transgene, which transgene was introduced into the animal or an ancestor of the animal at a prenatal, e.g., an embryonic stage. A transgene is a DNA which is integrated into the genome of a cell from which a transgenic animal develops. In one embodiment, cDNA encoding PRO295 polypeptide can be used to clone genomic DNA encoding PRO295 in accordance with established techniques and the genomic sequences used to generate transgenic animals that contain cells which express DNA encoding PRO295. Methods for generating transgenic animals, particularly animals such as mice or rats, have become conventional in the art and are described, for example, in U.S. Patent Nos. 4,736,866 and 4,870,009. Typically, particular cells would be targeted for PRO295 transgene incorporation with tissue-specific enhancers. Transgenic animals that include a copy of a transgene encoding PRO295 introduced into the germ line of the animal at an embryonic stage can be used to examine the effect of increased expression of DNA encoding PRO295. Such animals can be used as tester animals for reagents thought to confer protection from, for example, pathological conditions associated with its overexpression. In accordance with this facet of the invention, an animal is treated with the reagent and a reduced incidence of the pathological condition, compared to untreated animals bearing the transgene, would indicate a potential therapeutic intervention for the pathological condition.

Alternatively, non-human homologues of PRO295 can be used to construct a PRO295 "knock out" animal which has a defective or altered gene encoding PRO295 as a result of homologous recombination between the endogenous gene encoding PRO295 and altered genomic DNA encoding PRO295 introduced into an embryonic cell of the animal. For example, cDNA encoding PRO295 can be used to clone genomic DNA encoding PRO295 in accordance with established techniques. A portion of the genomic DNA encoding PRO295 can be deleted or replaced with another gene, such as a gene encoding a selectable marker which can be used to monitor integration. Typically, several kilobases of unaltered flanking DNA (both at the 5' and 3' ends) are included in the vector [see e.g., Thomas and Capecchi, Cell, 51:503 (1987) for a description of homologous recombination vectors]. The vector is introduced into an embryonic stem cell line (e.g., by electroporation) and cells in which the introduced DNA has homologously recombined with the endogenous DNA are selected [see e.g., Li et al., Cell, 69:915 (1992)]. The selected cells are then injected into a blastocyst of an animal (e.g., a mouse or rat) to form aggregation chimeras [see e.g., Bradley, in *Teratocarcinomas and Embryonic Stem Cells: A Practical Approach*, E. J. Robertson, ed. (IRL, Oxford, 1987), pp. 113-152]. A chimeric embryo can then be implanted into a suitable pseudopregnant female foster animal and the embryo brought to term to create a "knock out" animal. Progeny harboring the homologously recombined DNA in their germ cells can be identified by standard techniques and used to breed animals in which all cells of the animal contain the homologously recombined DNA. Knockout animals can be characterized for instance, for their ability to defend against certain pathological conditions and for their development of pathological conditions due to absence of the PRO295 polypeptide.

PRO295 polypeptides and portions thereof which have homology to integrin may also be useful for *in vivo* therapeutic purposes, as well as for various other applications. The identification of novel integrins and integrin-like molecules may have relevance to a number of human disorders such as modulating the binding or activity of cells of the immune system. Thus, the identification of new integrins and integrin-like molecules is of special importance in that such proteins may serve as potential therapeutics for a variety of different human disorders. Such polypeptides may also play important roles in biotechnological and medical research as well as various industrial applications. As a result, there is particular scientific and medical interest in new molecules, such as PRO 295.

E. Anti-PRO295 Antibodies

The present invention further provides anti-PRO295 polypeptide antibodies. Exemplary antibodies include polyclonal, monoclonal, humanized, bispecific, and heteroconjugate antibodies.

1. Polyclonal Antibodies

The anti-PRO295 antibodies of the present invention may comprise polyclonal antibodies. Methods of preparing polyclonal antibodies are known to the skilled artisan. Polyclonal antibodies can be raised in a mammal, for example, by one or more injections of an immunizing agent and, if desired, an adjuvant. Typically, the immunizing agent and/or adjuvant will be injected in the mammal by multiple subcutaneous or intraperitoneal injections. The immunizing agent may include the PRO295 polypeptide or a fusion protein thereof. It may be useful to conjugate the immunizing agent to a protein known to be immunogenic in the mammal being immunized. Examples of such immunogenic proteins include but are not limited to keyhole limpet hemocyanin, serum albumin, bovine thyroglobulin, and soybean trypsin inhibitor. Examples of adjuvants which may be employed include Freund's complete adjuvant and MPL-TDM adjuvant (monophosphoryl Lipid A, synthetic trehalose dicorynomycolate). The immunization protocol may be selected by one skilled in the art without undue experimentation.

2. Monoclonal Antibodies

The anti-PRO295 antibodies may, alternatively, be monoclonal antibodies. Monoclonal antibodies may be prepared using hybridoma methods, such as those described by Kohler and Milstein, Nature, 256:495 (1975). In a hybridoma method, a mouse, hamster, or other appropriate host animal, is typically immunized with an immunizing agent to elicit lymphocytes that produce or are capable of producing antibodies that will specifically bind to the immunizing agent. Alternatively, the lymphocytes may be immunized *in vitro*.

The immunizing agent will typically include the PRO295 polypeptide or a fusion protein thereof. Generally, either peripheral blood lymphocytes ("PBLs") are used if cells of human origin are desired, or spleen cells or lymph node cells are used if non-human mammalian sources are desired. The lymphocytes are then fused with an immortalized cell

line using a suitable fusing agent, such as polyethylene glycol, to form a hybridoma cell [Goding, Monoclonal Antibodies: Principles and Practice, Academic Press, (1986) pp. 59-103]. Immortalized cell lines are usually transformed mammalian cells, particularly myeloma cells of rodent, bovine and human origin. Usually, rat or mouse myeloma cell lines are employed. The hybridoma cells may be cultured in a suitable culture medium that preferably contains one or more substances that inhibit the growth or survival of the unfused, immortalized cells. For example, if the parental cells lack the enzyme hypoxanthine guanine phosphoribosyl transferase (HGPRT or HPRT), the culture medium for the hybridomas typically will include hypoxanthine, aminopterin, and thymidine ("HAT medium"), which substances prevent the growth of HGPRT-deficient cells.

Preferred immortalized cell lines are those that fuse efficiently, support stable high level expression of antibody by the selected antibody-producing cells, and are sensitive to a medium such as HAT medium. More preferred immortalized cell lines are murine myeloma lines, which can be obtained, for instance, from the Salk Institute Cell Distribution Center, San Diego, California and the American Type Culture Collection, Rockville, Maryland. Human myeloma and mouse-human heteromyeloma cell lines also have been described for the production of human monoclonal antibodies [Kozbor, J. Immunol., 133:3001 (1984); Brodeur et al., Monoclonal Antibody Production Techniques and Applications, Marcel Dekker, Inc., New York, (1987) pp. 51-63].

The culture medium in which the hybridoma cells are cultured can then be assayed for the presence of monoclonal antibodies directed against a PRO295 polypeptide. Preferably, the binding specificity of monoclonal antibodies produced by the hybridoma cells is determined by immunoprecipitation or by an *in vitro* binding assay, such as radioimmunoassay (RIA) or enzyme-linked immunoabsorbent assay (ELISA). Such techniques and assays are known in the art. The binding affinity of the monoclonal antibody can, for example, be determined by the Scatchard analysis of Munson and Pollard, Anal. Biochem., 107:220 (1980).

After the desired hybridoma cells are identified, the clones may be subcloned by limiting dilution procedures and grown by standard methods [Goding, supra]. Suitable culture media for this purpose include, for example, Dulbecco's Modified Eagle's Medium and RPMI-1640 medium. Alternatively, the hybridoma cells may be grown *in vivo* as ascites in a mammal.

The monoclonal antibodies secreted by the subclones may be isolated or purified from the culture medium or ascites fluid by conventional immunoglobulin purification procedures such as, for example, protein A-Sepharose, hydroxylapatite chromatography, gel electrophoresis, dialysis, or affinity chromatography.

The monoclonal antibodies may also be made by recombinant DNA methods, such as those described in U.S. Patent No. 4,816,567. DNA encoding the monoclonal antibodies of the invention can be readily isolated and sequenced using conventional procedures (e.g., by using oligonucleotide probes that are capable of binding specifically to genes encoding the heavy and light chains of murine antibodies). The hybridoma cells of the invention serve as a preferred source of such DNA. Once isolated, the DNA may be placed into expression vectors, which are then transfected into host cells such as simian COS cells, Chinese hamster ovary (CHO) cells, or myeloma cells that do not otherwise produce immunoglobulin protein, to obtain the synthesis of monoclonal antibodies in the recombinant host cells. The DNA also may be modified, for example, by substituting the coding sequence for human heavy and light chain constant domains in place of the homologous murine sequences [U.S. Patent No. 4,816,567; Morrison et al., supra] or by covalently joining to the immunoglobulin coding sequence all or part of the coding sequence for a non-immunoglobulin polypeptide. Such a non-immunoglobulin polypeptide can be substituted for the constant domains of an antibody of the invention, or can be substituted for the variable domains of one antigen-combining site of an antibody of the invention to create a chimeric bivalent antibody.

The antibodies may be monovalent antibodies. Methods for preparing monovalent antibodies are well known in the art. For example, one method involves recombinant expression of immunoglobulin light chain and modified heavy chain. The heavy chain is truncated generally at any point in the Fc region so as to prevent heavy chain crosslinking. Alternatively, the relevant cysteine residues are substituted with another amino acid residue or are deleted so as to prevent crosslinking.

In vitro methods are also suitable for preparing monovalent antibodies. Digestion of antibodies to produce fragments thereof, particularly, Fab fragments, can be accomplished using routine techniques known in the art.

3. Humanized Antibodies

The anti-PRO295 antibodies of the invention may further comprise humanized antibodies or human antibodies. Humanized forms of non-human (e.g., murine) antibodies are chimeric immunoglobulins, immunoglobulin chains or fragments thereof (such as Fv, Fab, Fab', $F(ab')_2$ or other antigen-binding subsequences of antibodies) which contain minimal sequence derived from non-human immunoglobulin. Humanized antibodies include human immunoglobulins (recipient antibody) in which residues from a complementary determining region (CDR) of the recipient are replaced by residues from a CDR of a non-human species (donor antibody) such as mouse, rat or rabbit having the desired specificity, affinity and capacity. In some instances, Fv framework residues of the human immunoglobulin are replaced by corresponding non-human residues. Humanized antibodies may also comprise residues which are found neither in the recipient antibody nor in the imported CDR or framework sequences. In general, the humanized antibody will comprise substantially all of at least one, and typically two, variable domains, in which all or substantially all of the CDR regions correspond to those of a non-human immunoglobulin and all or substantially all of the FR regions are those of a human immunoglobulin consensus sequence. The humanized antibody optimally also will comprise at least a portion of an immunoglobulin constant region (Fc), typically that of a human immunoglobulin [Jones et al., Nature, 321:522-525 (1986); Riechmann et al., Nature, 332:323-329 (1988); and Presta, Curr. Op. Struct. Biol., 2:593-596 (1992)].

Methods for humanizing non-human antibodies are well known in the art. Generally, a humanized antibody has one or more amino acid residues introduced into it from a source which is non-human. These non-human amino acid residues are often referred to as "import" residues, which are typically taken from an "import" variable domain. Humanization can be essentially performed following the method of Winter and co-workers [Jones et al., Nature, 321:522-525 (1986); Riechmann et al., Nature, 332:323-327 (1988); Verhoeyen et al., Science, 239:1534-1536 (1988)], by substituting rodent CDRs or CDR sequences for the corresponding sequences of a human antibody. Accordingly, such "humanized" antibodies are chimeric antibodies (U.S. Patent No. 4,816,567), wherein substantially less than an intact human variable domain has been substituted by the corresponding sequence from a non-human species. In practice, humanized antibodies are typically human antibodies in which some CDR

residues and possibly some FR residues are substituted by residues from analogous sites in rodent antibodies.

Human antibodies can also be produced using various techniques known in the art, including phage display libraries [Hoogenboom and Winter, J. Mol. Biol., 227:381 (1991); Marks et al., J. Mol. Biol., 222:581 (1991)]. The techniques of Cole et al. and Boerner et al. are also available for the preparation of human monoclonal antibodies (Cole et al., Monoclonal Antibodies and Cancer Therapy, Alan R. Liss, p. 77 (1985) and Boerner et al., J. Immunol., 147(1):86-95 (1991)].

4. Bispecific Antibodies

Bispecific antibodies are monoclonal, preferably human or humanized, antibodies that have binding specificities for at least two different antigens. In the present case, one of the binding specificities is for a PRO295 polypeptide, the other one is for any other antigen, and preferably for a cell-surface protein or receptor or receptor subunit.

Methods for making bispecific antibodies are known in the art. Traditionally, the recombinant production of bispecific antibodies is based on the co-expression of two immunoglobulin heavy-chain/light-chain pairs, where the two heavy chains have different specificities [Milstein and Cuello, Nature, 305:537-539 (1983)]. Because of the random assortment of immunoglobulin heavy and light chains, these hybridomas (quadromas) produce a potential mixture of ten different antibody molecules, of which only one has the correct bispecific structure. The purification of the correct molecule is usually accomplished by affinity chromatography steps. Similar procedures are disclosed in WO 93/08829, published 13 May 1993, and in Traunecker et al., EMBO J., 10:3655-3659 (1991).

Antibody variable domains with the desired binding specificities (antibody-antigen combining sites) can be fused to immunoglobulin constant domain sequences. The fusion preferably is with an immunoglobulin heavy-chain constant domain, comprising at least part of the hinge, CH2, and CH3 regions. It is preferred to have the first heavy-chain constant region (CH1) containing the site necessary for light-chain binding present in at least one of the fusions. DNAs encoding the immunoglobulin heavy-chain fusions and, if desired, the immunoglobulin light chain, are inserted into separate expression vectors, and are co-transfected into a suitable host organism. For further details of generating bispecific

antibodies see, for example, Suresh et al., Methods in Enzymology, 121:210 (1986).

5. Heteroconjugate Antibodies

Heteroconjugate antibodies are also within the scope of the present invention. Heteroconjugate antibodies are composed of two covalently joined antibodies. Such antibodies have, for example, been proposed to target immune system cells to unwanted cells [U.S. Patent No. 4,676,980], and for treatment of HIV infection [WO 91/00360; WO 92/200373; EP 03089]. It is contemplated that the antibodies may be prepared *in vitro* using known methods in synthetic protein chemistry, including those involving crosslinking agents. For example, immunotoxins may be constructed using a disulfide exchange reaction or by forming a thioether bond. Examples of suitable reagents for this purpose include iminothiolate and methyl-4-mercaptobutyrimidate and those disclosed, for example, in U.S. Patent No. 4,676,980.

F. Uses for anti-PRO295 Antibodies

The anti-PRO295 antibodies of the present invention have various utilities. For example, anti-PRO295 antibodies may be used in diagnostic assays for PRO295 polypeptides, *e.g.*, detecting expression in specific cells, tissues, or serum. Various diagnostic assay techniques known in the art may be used, such as competitive binding assays, direct or indirect sandwich assays and immunoprecipitation assays conducted in either heterogeneous or homogeneous phases [Zola, Monoclonal Antibodies: A Manual of Techniques, CRC Press, Inc. (1987) pp. 147-158]. The antibodies used in the diagnostic assays can be labeled with a detectable moiety. The detectable moiety should be capable of producing, either directly or indirectly, a detectable signal. For example, the detectable moiety may be a radioisotope, such as ^{3}H, ^{14}C, ^{32}P, ^{35}S, or ^{125}I, a fluorescent or chemiluminescent compound, such as fluorescein isothiocyanate, rhodamine, or luciferin, or an enzyme, such as alkaline phosphatase, beta-galactosidase or horseradish peroxidase. Any method known in the art for conjugating the antibody to the detectable moiety may be employed, including those methods described by Hunter et al., Nature, 144:945 (1962); David et al., Biochemistry, 13:1014 (1974); Pain et al., J. Immunol. Meth., 40:219 (1981); and Nygren, J. Histochem. and Cytochem., 30:407 (1982).

Anti-PRO295 antibodies also are useful for the affinity purification of PRO295 polypeptides from recombinant cell culture or natural sources. In this process, the antibodies against a PRO295 polypeptide are immobilized on a suitable support, such a Sephadex resin or filter paper, using methods well known in the art. The immobilized antibody then is contacted with a sample containing the PRO295 polypeptide to be purified, and thereafter the support is washed with a suitable solvent that will remove substantially all the material in the sample except the PRO295 polypeptide, which is bound to the immobilized antibody. Finally, the support is washed with another suitable solvent that will release the PRO295 polypeptide from the antibody.

The following examples are offered for illustrative purposes only, and are not intended to limit the scope of the present invention in any way.

All patent and literature references cited in the present specification are hereby incorporated by reference in their entirety.

EXAMPLES

Commercially available reagents referred to in the examples were used according to manufacturer's instructions unless otherwise indicated. The source of those cells identified in the following examples, and throughout the specification, by ATCC accession numbers is the American Type Culture Collection, Rockville, Maryland.

EXAMPLE 1: Isolation of cDNA Clones Encoding Human PRO295

The extracellular domain (ECD) sequences (including the secretion signal, if any) of from about 950 known secreted proteins from the Swiss-Prot public protein database were used to search expressed sequence tag (EST) databases. The EST databases included public EST databases (e.g., GenBank) and a proprietary EST DNA database (LIFESEQ™, Incyte Pharmaceuticals, Palo Alto, CA). The search was performed using the computer program BLAST or BLAST2 (Altshul et al., Methods in Enzymology 266:460-480 (1996)) as a comparison of the ECD protein sequences to a 6 frame translation of the EST sequence. Those comparisons resulting in a BLAST score of 70 (or in some cases 90) or greater that did not encode known proteins were clustered and assembled into consensus DNA sequences with the program "phrap" (Phil Green, University of Washington, Seattle, Washington;

http://bozeman.mbt.washington.edu/phrap.docs/phrap.html).

A consensus DNA sequence was assembled relative to other EST sequences using phrap. This consensus sequence is herein designated DNA35814 (see Figures 3A and 3B). In some cases, the consensus DNA sequence was extended using repeated cycles of BLAST and phrap to extend the consensus sequence as far as possible using the sources of EST sequences discussed above. The extended assembly sequence, is indicated as a number of second alignment figures, as shown in Figures 4A-AB.

Based on the DNA35814 consensus sequence, oligonucleotides were synthesized: 1) to identify by PCR a cDNA library that contained the sequence of interest, and 2) for use as probes to isolate a clone of the full-length coding sequence for PRO295. Forward and reverse PCR primers generally range from 20 to 30 nucleotides and are often designed to give a PCR product of about 100-1000 bp in length. The probe sequences are typically 40-55 bp in length. In some cases, additional oligonucleotides are synthesized when the consensus sequence is greater than about 1-1.5kbp. In order to screen several libraries for a full-length clone, DNA from the libraries was screened by PCR amplification, as ber Ausubel et al., Current Protocols in Molecular Biology, with the PCR primer pair. A positive library was then used to isolate clones encoding the gene of interest by the *in vivo* cloning procedure using the probe oligonucleotide and one of the primer pairs.

Forward and reverse PCR primers were synthesized:

forward PCR primer (.f1)	5'-GCAGAGCGGAGATGCAGCGGCTTG-3'	(SEQ ID NO:4)
forward PCR primer (.f2)	5'-CCCAGCATGTACTGCCAG-3'	(SEQ ID NO:5)
forward PCR primer (.f3)	5'-TTGGCAGCTTCATGGAGG-3'	(SEQ ID NO:6)
forward PCR primer (.f4)	5'-CCTGGGCAAAAATGCAAC-3'	(SEQ ID NO:7)

reverse PCR primer (.r1) 5'-CTCCAGCTCCTGGCGCACCTCCTC-3' (SEQ ID NO:8)

Additionally, a synthetic oligonucleotide hybridization probe was constructed from the consensus DNA35814 sequence which had the following nucleotide sequence

hybridization probe

5'-GGCTCTCAGCTACCGCGCAGGAGCGAGGCCACCCTCAATGAGATG-3'

(SEQ ID NO:10)

In order to screen several libraries for a source of a full-length clone, DNA from the libraries was screened by PCR amplification with the PCR primer pairs identified above. A positive library was then used to isolate clones encoding the PRO295 gene using the probe oligonucleotide and one of the PCR primers.

RNA for construction of the cDNA libraries was isolated from human fetal lung tissue. The cDNA libraries used to isolate the cDNA clones were constructed by standard methods using commercially available reagents such as those from Invitrogen, San Diego, CA. The cDNA was primed with oligo dT containing a NotI site, linked with blunt to SalI hemikinased adaptors, cleaved with NotI, sized appropriately by gel electrophoresis, and cloned in a defined orientation into a suitable cloning vector (such as pRKB or pRKD; pRK5B is a precursor of pRK5D that does not contain the SfiI site; see, Holmes et al., Science, 253:1278-1280 (1991)) in the unique XhoI and NotI sites.

DNA sequencing of the clones isolated as described above gave the full-length DNA sequence for PRO295 [herein designated as UNQ258 (DNA38268)] (SEQ ID NO:1) and the derived protein sequence for PRO295.

The entire nucleotide sequence of UNQ258 (DNA38268) is shown in Figure 1 (SEQ ID NO:1). Clone UNQ258 (DNA38268) contains a single open reading frame with an apparent translational initiation site at nucleotide positions 153-155 [Kozak et al., supra] and ending at the stop codon at nucleotide positions 1202-1205 (Figure 1). The predicted polypeptide precursor is 350 amino acids long (Figure 3). Clone UNQ258 (DNA38268) has been deposited with ATCC and is assigned ATCC deposit no. NEED DEPOSIT INFORMATION.

Analysis of the amino acid sequence of the full-length PRO295 polypeptide suggests that portions of it possess significant homology to the integrin proteins as shown in Figures 5 and 6, thereby indicating that PRO295 may be a novel integrin.

EXAMPLE 2: Use of PRO295-encoding DNA as a hybridization probe

The following method describes use of a nucleotide sequence encoding PRO295 as a hybridization probe.

DNA comprising the coding sequence of full-length PRO295 (as shown in Figure 1, SEQ ID NO:1) is employed as a probe to screen for homologous DNAs (such as those encoding naturally-occurring variants of PRO295) in human tissue cDNA libraries or human tissue genomic libraries.

Hybridization and washing of filters containing either library DNAs is performed under the following high stringency conditions. Hybridization of radiolabeled UNQ258 (DNA38268)-derived probe to the filters is performed in a solution of 50% formamide, 5x SSC, 0.1% SDS, 0.1% sodium pyrophosphate, 50 mM sodium phosphate, pH 6.8, 2x Denhardt's solution, and 10% dextran sulfate at 42°C for 20 hours. Washing of the filters is performed in an aqueous solution of 0.1x SSC and 0.1% SDS at 42°C.

DNAs having a desired sequence identity with the DNA encoding full-length native sequence PRO295 polypeptide can then be identified using standard techniques known in the art.

EXAMPLE 3: Expression of PRO295 Polypeptides in *E. coli*

This example illustrates the preparation of unglycosylated forms of PRO295 polypeptides by recombinant expression in *E. coli*.

The DNA sequence encoding PRO295 (SEQ ID NO:3) is initially amplified using selected PCR primers. The primers should contain restriction enzyme sites which correspond to the restriction enzyme sites on the selected expression vector. A variety of expression vectors may be employed. An example of a suitable vector is pBR322 (derived from *E. coli*; see Bolivar et al., Gene, 2:95 (1977)) which contains genes for ampicillin and tetracycline resistance. The vector is digested with restriction enzyme and dephosphorylated. The PCR amplified sequences are then ligated into the vector. The vector will preferably include sequences which encode for an antibiotic resistance gene, a trp promoter, a polyhis leader (including the first six STII codons, polyhis sequence, and enterokinase cleavage site), the PRO295 coding region, lambda transcriptional terminator, and an argU gene.

The ligation mixture is then used to transform a selected *E. coli* strain using the

methods described in Sambrook et al., supra. Transformants are identified by their ability to grow on LB plates and antibiotic resistant colonies are then selected. Plasmid DNA can be isolated and confirmed by restriction analysis and DNA sequencing.

Selected clones can be grown overnight in liquid culture medium such as LB broth supplemented with antibiotics. The overnight culture may subsequently be used to inoculate a larger scale culture. The cells are then grown to a desired optical density, during which the expression promoter is turned on.

After culturing the cells for several more hours, the cells can be harvested by centrifugation. The cell pellet obtained by the centrifugation can be solubilized using various agents known in the art, and the solubilized PRO295 polypeptide can then be purified using a metal chelating column under conditions that allow tight binding of the polypeptide.

EXAMPLE 4: Expression of PRO295 Polypeptides in Mammalian Cells

This example illustrates preparation of glycosylated forms of PRO295 polypeptides by recombinant expression in mammalian cells.

The vector, pRK5 (see EP 307,247, published March 15, 1989), is employed as the expression vector. Optionally, the PRO295-encoding DNA is ligated into pRK5 with selected restriction enzymes to allow insertion of the PRO295-encoding DNA using ligation methods such as described in Sambrook et al., supra. The resulting vector is called pRK5-PRO295

In one embodiment, the selected host cells may be 293 cells. Human 293 cells (ATCC CCL 1573) are grown to confluence in tissue culture plates in medium such as DMEM supplemented with fetal calf serum and optionally, nutrient components and/or antibiotics. About 10 μg pRK5-PRO295 DNA is mixed with about 1 μg DNA encoding the VA RNA gene [Thimmappaya et al., Cell, 31:543 (1982)] and dissolved in 500 μl of 1 mM Tris-HCl, 0.1 mM EDTA, 0.227 M $CaCl_2$. To this mixture is added, dropwise, 500 μl of 50 mM HEPES (pH 7.35), 280 mM NaCl, 1.5 mM $NaPO_4$, and a precipitate is allowed to form for 10 minutes at 25°C. The precipitate is suspended and added to the 293 cells and allowed to settle for about four hours at 37°C. The culture medium is aspirated off and 2 ml of 20% glycerol in PBS is added for 30 seconds. The 293 cells are then washed with serum free medium, fresh medium is added and the cells are incubated for about 5 days.

Approximately 24 hours after the transfections, the culture medium is removed and

replaced with culture medium (alone) or culture medium containing 200 μCi/ml ^{35}S-cysteine and 200 μCi/ml ^{35}S-methionine. After a 12 hour incubation, the conditioned medium is collected, concentrated on a spin filter, and loaded onto a 15% SDS gel. The processed gel may be dried and exposed to film for a selected period of time to reveal the presence of PRO295 polypeptide. The cultures containing transfected cells may undergo further incubation (in serum free medium) and the medium is tested in selected bioassays.

In an alternative technique, PRO295-encoding DNA may be introduced into 293 cells transiently using the dextran sulfate method described by Somparyrac et al., Proc. Natl. Acad. Sci., 12:7575 (1981). 293 cells are grown to maximal density in a spinner flask and 700 μg pRK5-PRO295 DNA is added. The cells are first concentrated from the spinner flask by centrifugation and washed with PBS. The DNA-dextran precipitate is incubated on the cell pellet for four hours. The cells are treated with 20% glycerol for 90 seconds, washed with tissue culture medium, and re-introduced into the spinner flask containing tissue culture medium, 5 μg/ml bovine insulin and 0.1 μg/ml bovine transferrin. After about four days, the conditioned media is centrifuged and filtered to remove cells and debris. The sample containing expressed PRO295 polypeptide can then be concentrated and purified by any selected method, such as dialysis and/or column chromatography.

In another embodiment, PRO295 polypeptide can be expressed in CHO cells. The pRK5-PRO295 vector can be transfected into CHO cells using known reagents such as $CaPO_4$ or DEAE-dextran. As described above, the cell cultures can be incubated, and the medium replaced with culture medium (alone) or medium containing a radiolabel such as ^{35}S-methionine. After determining the presence of PRO295 polypeptide, the culture medium may be replaced with serum free medium. Preferably, the cultures are incubated for about 6 days, and then the conditioned medium is harvested. The medium containing the expressed PRO295 polypeptide can then be concentrated and purified by any selected method.

Epitope-tagged PRO295 polypeptide may also be expressed in host CHO cells. The PRO295-encoding DNA may be subcloned out of the pRK5 vector. The subclone insert can undergo PCR to fuse in frame with a selected epitope tag such as a poly-his tag into a Baculovirus expression vector. The poly-his tagged PRO295-encoding DNA insert can then be subcloned into a SV40 driven vector containing a selection marker such as DHFR for selection of stable clones. Finally, the CHO cells can be transfected (as described above) with

the SV40 driven vector. Labeling may be performed, as described above, to verify expression. The culture medium containing the expressed poly-His tagged PRO295 polypeptide can then be concentrated and purified by any selected method, such as by Ni^{2+}-chelate affinity chromatography.

EXAMPLE 5: Expression of a PRO295 Polypeptide in Yeast

The following method describes recombinant expression of PRO295 polypeptides in yeast.

First, yeast expression vectors are constructed for intracellular production or secretion of PRO295 polypeptide from the ADH2/GAPDH promoter. DNA encoding the PRO295 polypeptide of interest, a selected signal peptide and the promoter is inserted into suitable restriction enzyme sites in the selected plasmid to direct intracellular expression of the PRO295 polypeptide. For secretion, DNA encoding the PRO295 polypeptide can be cloned into the selected plasmid, together with DNA encoding the ADH2/GAPDH promoter, the yeast alpha-factor secretory signal/leader sequence, and linker sequences (if needed) for expression of the PRO295 polypeptide.

Yeast cells, such as yeast strain AB110, can then be transformed with the expression plasmids described above and cultured in selected fermentation media. The transformed yeast supernatants can be analyzed by precipitation with 10% trichloroacetic acid and separation by SDS-PAGE, followed by staining of the gels with Coomassie Blue stain.

Recombinant PRO295 polypeptide can subsequently be isolated and purified by removing the yeast cells from the fermentation medium by centrifugation and then concentrating the medium using selected cartridge filters. The concentrate containing the PRO295 polypeptide may further be purified using selected column chromatography resins.

EXAMPLE 6: Expression of PRO295 Polypeptides in Baculovirus

The following method describes recombinant expression of PRO295 polypeptides in Baculovirus.

The PRO295-encoding DNA is fused upstream of an epitope tag contained within a baculovirus expression vector. Such epitope tags include poly-his tags and immunoglobulin tags (like Fc regions of IgG). A variety of plasmids may be employed, including plasmids

derived from commercially available plasmids such as pVL1393 (Novagen). Briefly, the PRO295-encoding DNA or the desired portion of the PRO295-encoding DNA (such as the sequence encoding the extracellular domain of a transmembrane protein) is amplified by PCR with primers complementary to the 5' and 3' regions. The 5' primer may incorporate flanking (selected) restriction enzyme sites. The product is then digested with those selected restriction enzymes and subcloned into the expression vector.

Recombinant baculovirus is generated by co-transfecting the above plasmid and BaculoGold™ virus DNA (Pharmingen) into *Spodoptera frugiperda* ("Sf9") cells (ATCC CRL 1711) using lipofectin (commercially available from GIBCO-BRL). After 4 to 5 days of incubation at 28°C, the released viruses are harvested and used for further amplifications. Viral infection and protein expression is performed as described by O'Reilley et al., Baculovirus expression vectors: A laboratory Manual, Oxford:Oxford University Press (1994).

Expressed poly-his tagged PRO295 polypeptide can then be purified, for example, by Ni^{2+}-chelate affinity chromatography as follows. Extracts are prepared from recombinant virus-infected Sf9 cells as described by Rupert et al., Nature, 362:175-179 (1993). Briefly, Sf9 cells are washed, resuspended in sonication buffer (25 mL Hepes, pH 7.9; 12.5 mM $MgCl_2$; 0.1 mM EDTA; 10% Glycerol; 0.1% NP-40; 0.4 M KCl), and sonicated twice for 20 seconds on ice. The sonicates are cleared by centrifugation, and the supernatant is diluted 50-fold in loading buffer (50 mM phosphate, 300 mM NaCl, 10% Glycerol, pH 7.8) and filtered through a 0.45 μm filter. A Ni^{2+}-NTA agarose column (commercially available from Qiagen) is prepared with a bed volume of 5 mL, washed with 25 mL of water and equilibrated with 25 mL of loading buffer. The filtered cell extract is loaded onto the column at 0.5 mL per minute. The column is washed to baseline A_{280} with loading buffer, at which point fraction collection is started. Next, the column is washed with a secondary wash buffer (50 mM phosphate; 300 mM NaCl, 10% Glycerol, pH 6.0), which elutes nonspecifically bound protein. After reaching A_{280} baseline again, the column is developed with a 0 to 500 mM Imidazole gradient in the secondary wash buffer. One mL fractions are collected and analyzed by SDS-PAGE and silver staining or western blot with Ni^{2+}-NTA-conjugated to alkaline phosphatase (Qiagen). Fractions containing the eluted His_{10}-tagged PRO295 polypeptide are pooled and dialyzed against loading buffer.

Alternatively, purification of the IgG tagged (or Fc tagged) PRO295 polypeptide can

be performed using known chromatography techniques, including for instance, Protein A or protein G column chromatography.

EXAMPLE 7: Preparation of Antibodies that Bind PRO295 Polypeptides

This example illustrates the preparation of monoclonal antibodies which can specifically bind to PRO295 polypeptides.

Techniques for producing the monoclonal antibodies are known in the art and are described, for instance, in Goding, supra. Immunogens that may be employed include purified PRO295 polypeptide, fusion proteins containing a PRO295 polypeptide, and cells expressing recombinant PRO295 polypeptide on the cell surface. Selection of the immunogen can be made by the skilled artisan without undue experimentation.

Mice, such as Balb/c, are immunized with the PRO295 immunogen emulsified in complete Freund's adjuvant and injected subcutaneously or intraperitoneally in an amount from 1-100 micrograms. Alternatively, the immunogen is emulsified in MPL-TDM adjuvant (Ribi Immunochemical Research, Hamilton, MT) and injected into the animal's hind foot pads. The immunized mice are then boosted 10 to 12 days later with additional immunogen emulsified in the selected adjuvant. Thereafter, for several weeks, the mice may also be boosted with additional immunization injections. Serum samples may be periodically obtained from the mice by retro-orbital bleeding for testing in ELISA assays to detect anti-PRO295 polypeptide antibodies.

After a suitable antibody titer has been detected, the animals "positive" for antibodies can be injected with a final intravenous injection of PRO295 polypeptide. Three to four days later, the mice are sacrificed and the spleen cells are harvested. The spleen cells are then fused (using 35% polyethylene glycol) to a selected murine myeloma cell line such as P3X63AgU.1, available from ATCC, No. CRL 1597. The fusions generate hybridoma cells which can then be plated in 96 well tissue culture plates containing HAT (hypoxanthine, aminopterin, and thymidine) medium to inhibit proliferation of non-fused cells, myeloma hybrids, and spleen cell hybrids.

The hybridoma cells will be screened in an ELISA for reactivity against PRO295 polypeptide. Determination of "positive" hybridoma cells secreting the desired monoclonal antibodies against a PRO295 polypeptide is within the skill in the art.

The positive hybridoma cells can be injected intraperitoneally into syngeneic Balb/c mice to produce ascites containing the anti-PRO295 polypeptide monoclonal antibodies. Alternatively, the hybridoma cells can be grown in tissue culture flasks or roller bottles. Purification of the monoclonal antibodies produced in the ascites can be accomplished using ammonium sulfate precipitation, followed by gel exclusion chromatography. Alternatively, affinity chromatography based upon binding of antibody to protein A or protein G can be employed.

Deposit of Material

The following materials have been deposited with the American Type Culture Collection, 12301 Parklawn Drive, Rockville, MD, USA (ATCC):

Material	ATCC Dep. No.	Deposit Date

NEED DEPOSIT INFORMATION

This deposit was made under the provisions of the Budapest Treaty on the International Recognition of the Deposit of Microorganisms for the Purpose of Patent Procedure and the Regulations thereunder (Budapest Treaty). This assures maintenance of a viable culture of the deposit for 30 years from the date of deposit. The deposit will be made available by ATCC under the terms of the Budapest Treaty, and subject to an agreement between Genentech, Inc. and ATCC, which assures permanent and unrestricted availability of the progeny of the culture of the deposit to the public upon issuance of the pertinent U.S. patent or upon laying open to the public of any U.S. or foreign patent application, whichever comes first, and assures availability of the progeny to one determined by the U.S. Commissioner of Patents and Trademarks to be entitled thereto according to 35 USC §122 and the Commissioner's rules pursuant thereto (including 37 CFR §1.14 with particular reference to 886 OG 638).

The assignee of the present application has agreed that if a culture of the materials on deposit should die or be lost or destroyed when cultivated under suitable conditions, the materials will be promptly replaced on notification with another of the same. Availability of the deposited material is not to be construed as a license to practice the invention in contravention of the rights granted under the authority of any government in accordance with its patent laws.

The foregoing written specification is considered to be sufficient to enable one skilled in the art to practice the invention. The present invention is not to be limited in scope by the construct deposited, since the deposited embodiment is intended as a single illustration of certain aspects of the invention and any constructs that are functionally equivalent are within the scope of this invention. The deposit of material herein does not constitute an admission that the written description herein contained is inadequate to enable the practice of any aspect of the invention, including the best mode thereof, nor is it to be construed as limiting the scope of the claims to the specific illustrations that it represents. Indeed, various modifications of the invention in addition to those shown and described herein will become apparent to those skilled in the art from the foregoing description and fall within the scope of the appended claims.

WHAT IS CLAIMED IS:

1. Isolated nucleic acid comprising DNA encoding a PRO295 polypeptide having amino acid residues 1 to 350 of Figure 2 (SEQ ID NO:3).

2. The nucleic acid of Claim 1 wherein said DNA comprises the nucleotide sequence of SEQ ID NO:1 or its complement.

3. The nucleic acid of Claim 1 wherein said DNA comprises nucleotides 153-1202 of the nucleotide seqeunce of SEQ ID NO:1 (SEQ ID NO:2).

4. An isolated nucleic acid comprising the nucleotide sequence of the full-length coding sequence of clone UNQ258 (DNA38268) deposited under accession number ATCC NEED DEPOSIT INFORMATION.

5. A vector comprising the nucleic acid of any one of Claim 1 or Claim 4.

6. The vector of Claim 5 operably linked to control sequences recognized by a host cell transformed with the vector.

7. A host cell comprising the vector of Claim 5.

8. The host cell of Claim 7 wherein said cell is a CHO cell.

9. The host cell of Claim 7 wherein said cell is an *E. coli*.

10. The host cell of Claim 7 wherein said cell is a yeast cell.

11. A process for producing a PRO295 polypeptide comprising culturing the host cell of Claim 7 under conditions suitable for expression of said PRO295 polypeptide and recovering said PRO295 polypeptide from the cell culture.

12. Isolated native sequence PRO295 polypeptide comprising amino acid residues 1 to 350 of Figure 2 (SEQ ID NO:3).

13. A chimeric molecule comprising a PRO295 polypeptide fused to a heterologous amino acid sequence.

14. The chimeric molecule of Claim 13 wherein said heterologous amino acid sequence is an epitope tag sequence.

15. The chimeric molecule of Claim 13 wherein said heterologous amino acid sequence is a Fc region of an immunoglobulin.

16. An antibody which specifically binds to a PRO295 polypeptide.

17. The antibody of Claim 16 wherein said antibody is a monoclonal antibody.

Abstract of the Disclosure

The present invention is directed to novel polypeptides having homology to the integrin protein and to nucleic acid molecules encoding those polypeptides. Also provided herein are vectors and host cells comprising those nucleic acid sequences, chimeric polypeptide molecules comprising the polypeptides of the present invention fused to heterologous polypeptide sequences, antibodies which bind to the polypeptides of the present invention and to methods for producing the polypeptides of the present invention.

FIGURE 1A

CGCCGCGCTCCCGCACCCGCGGCCCGCCCACCGCGCCGCTCCCGCATCTGCACCCGCAGC
CCGGCGGCCTCCCGGCGGGAGCGAGCAGATCCAGTCCGGCCCGCAGCGCAACTCGGTCCA
GTCGGGGCGGCGGCTGCGGGCGCAGAGCGGAG
><MET {trans=1-s, dir=f, res=1}
ATGCAGCGGCTTGGGGCCACCCTGCTGTGCCTGCTGCTGGCGGCGGCGGTCCCCACGGCC
CCCGCGCCCGCTCCGACGGCGACCTCGGCTCCAGTCAAGCCCGGCCCGGCTCTCAGCTAC
CCGCAGGAGGAGGCCACCCTCAATGAGATGTTCCGCGAGGTTGAGGAACTGATGGAGGAC
ACGCAGCACAAATTGCGCAGCGCGGTGGAAGAGATGGAGGCAGAAGAAGCTGCTGCTAAA
GCATCATCAGAAGTGAACCTGGCAAACTTACCTCCCAGCTATCACAATGAGACCAACACA
GACACGAAGGTTGGAAATAATACCATCCATGTGCACCGAGAAATTCACAAGATAACCAAC
AACCAGACTGGACAAATGGTCTTTTCAGAGACAGTTATCACATCTGTGGGAGACGAAGAA
GGCAGAAGGAGCCACGAGTGCATCATCGACGAGGACTGTGGGCCCAGCATGTACTGCCAG
TTTGCCAGCTTCCAGTACACCTGCCAGCCATGCCGGGGCCAGAGGATGCTCTGCACCCGG
GACAGTGAGTGCTGTGGAGACCAGCTGTGTGTCTGGGGTCACTGCACCAAAATGGCCACC
AGGGGCAGCAATGGGACCATCTGTGACAACCAGAGGGACTGCCAGCCGGGGCTGTGCTGT
GCCTTCCAGAGAGGCCTGCTGTTCCCTGTGTGCACACCCCTGCCCGTGGAGGGCGAGCTT
TGCCATGACCCCGCCAGCCGGCTTCTGGACCTCATCACCTGGGAGCTAGAGCCTGATGGA
GCCTTGGACCGATGCCCTTGTGCCAGTGGCCTCCTCTGCCAGCCCCACAGCCACAGCCTG
GTGTATGTGTGCAAGCCGACCTTCGTGGGGAGCCGTGACCAAGATGGGGAGATCCTGCTG
CCCAGAGAGGTCCCCGATGAGTATGAAGTTGGCAGCTTCATGGAGGAGGTGCGCCAGGAG
CTGGAGGACCTGGAGAGGAGCCTGACTGAAGAGATGGCGCTGGGGGAGCCTGCGGCTGCC
GCCGCTGCACTGCTGGGAGGGGAAGAGATTTAGATCTGGACCAGGCTGTGGGTAGATGTG
CAATAGAAATAGCTAATTTATTTCCCCAGGTGTGTGCTTTAGGCGTGGGCTGACCAGGCT
TCTTCCTACATCTTCTTCCCAGTAAGTTTCCCCTCTGGCTTGACAGCATGAGGTGTTGTG
CATTTGTTCAGCTCCCCCAGGCTGTTCTCCAGGCTTCACAGTCTGGTGCTTGGGAGAGTC
AGGCAGGGTTAAACTGCAGGAGCAGTTTGCCACCCCTGTCCAGATTATTGGCTGCTTTGC
CTCTACCAGTTGGCAGACAGCCGTTTGTTCTACATGGCTTTGATAATTGTTTGAGGGGAG
GAGATGGAAACAATGTGGAGTCTCCCTCTGATTGGTTTTGGGGAAATGTGGAGAAGAGTG
CCCTGCTTTGCAAACATCAACCTGGCAAAAATGCAACAAATGAATTTTCCACGCAGTTCT
TTCCATGGGCATAGGTAAGCTGTGCCTTCAGCTGTTGCAGATGAAATGTTCTGTTCACCC
TGCATTACATGTGTTTATTCATCCAGCAGTGTTGCTCAGCTCCTACCTCTGTGCCAGGGC
AGCATTTTCATATCCAAGATCAATTCCCTCTCTCAGCACAGCCTGGGGAGGGGGTCATTG
TTCTCCTCGTCCATCAGGGATCTCAGAGGCTCAGAGACTGCAAGCTGCTTGCCCAAGTCA
CACAGCTAGTGAAGACCAGAGCAGTTTCATCTGGTTGTGACTCTAAGCTCAGTGCTCTCT
CCACTACCCCACACCAGCCTTGGTGCCACCAAAAGTGCTCCCCAAAAGGAAGGAGAATGG

FIGURE 1B

GATTTTTCTTGAGGCATGCACATCTGGAATTAAGGTCAAACTAATTCTCACATCCCTCTA
AAAGTAAACTACTGTTAGGAACAGCAGTGTTCTCACAGTGTGGGGCAGCCGTCCTTCTAA
TGAAGACAATGATATTGACACTGTCCCTCTTTGGCAGTTGCATTAGTAACTTTGAAAGGT
ATATGACTGAGCGTAGCATACAGGTTAACCTGCAGAAACAGTACTTAGGTAATTGTAGGG
CGAGGATTATAAATGAAATTTGCAAAATCACTTAGCAGCAACTGAAGACAATTATCAACC
ACGTGGAGAAAATCAAACCGAGCAGGGCTGTGTGAAACATGGTTGTAATATGCGACTGCG
AACACTGAACTCTACGCCACTCCACAAATGATGTTTTCAGGTGTCATGGACTGTTGCCAC
CATGTATTCATCCAGAGTTCTTAAAGTTTAAAGTTGCACATGATTGTATAAGCATGCTTT
CTTTGAGTTTTAAATTATGTATAAACATAAGTTGCATTTAGAAATCAAGCATAAATCACT
TCAACTGCAAAAAAAAAAAAAAAAAAAAAAAAAA

FIGURE 2

MQRLGATLLCLLLAAAVPTAPAPAPTATSAPVKPGPALSYPQEEATLNEMFREVEELMED
TQHKLRSAVEEMEAEEAAAKASSEVNLANLPPSYHNETNTDTKVGNNTIHVHREIHKITN
NQTGQMVFSETVITSVGDEEGRRSHECIIDEDCGPSMYCQFASFQYTCQPCRGQRMLCTR
DSECCGDQLCVWGHCTKMATRGSNGTICDNQRDCQPGLCCAFQRGLLFPVCTPLPVEGEL
CHDPASRLLDLITWELEPDGALDRCPCASGLLCQPHSHSLVYVCKPTFVGSRDQDGEILL
PREVPDEYEVGSFMEEVRQELEDLERSLTEEMALGEPAAAAAALLGGEEI

FIGURE 3A

```
2054176          1 CTGTGGGAGACGAAGAAGGCAGAAGGAGCCACGAGTGCATCATCGACGAG
1756022          1   TGTGGGAGACGAAGAAGGCAGAAGGAGCCACGAGTGCATCATCGACGAG
980963           1          GACGAAGAAGGCAGAAGGAGCCACGAGTGCATCATCGACGAG
W61032           1                      GCCACGAGTGCATCATCGACGAG
<DNA35814:       1 CTGTGGGAGACGAAGAAGGCAGAAGGAGCCACGAGTGCATCATCGACGAG

2054176         51 GACTGTGTGGGCCCAGCATGTACTGCCAGTTTGCCAGCTTCCAGTACACCTG
1756022         50 GACTGTGTGGGCCCAGCATGTACTGCCAGTTTGCCAGCTTCCAGTACACCTG
980963          43 GACTGTGTGGGCCCAGCATGTACTGCCAGTTTGCCAGCTTCCAGTACACCTG
W61032          24 GACTGTGTGGGCCCAGCATGTACTGCCAGTTTGCCAGCTTCCAGTACACCTG
853079           1                                      AGTACACCTG
<DNA35814:      51 GACTGTGTGGGCCCAGCATGTACTGCCAGTTTGCCAGCTTCCAGTACACCTG

2054176        101 CCAGCCATGCCGGGGCCAGAGGATGCTCTGCACCCGGGACAGTNAGTGCT
1756022        100 CCAGCCATGCCGGGGCCAGAGGATGCTCTGCACCCGGGACAGTGAGTGCT
980963          93 CCAGCCATGCCGGGGCCAGAGGATGCTCTGCACCCGGGACAGTGAGTGCT
W61032          74 CCAGCCATGCCGGGGCCAGAGGATGCTCTGCACCCGGGACAGTGAGTGCT
853079          11 CCAGCCATGCCGGGGCCAGAGGATGCTCTGCACCCGGGACAGTGAGTGCT
<DNA35814:     101 CCAGCCATGCCGGGGCCAGAGGATGCTCTGCACCCGGGACAGTGAGTGCT

2054176        151 GTGGAGACCAGCTGTGTGTCTGGGGTCACTGCACCAAAATGGCCACCAGG
1756022        150 GTGGAGACCAGCTGTGTGTCTGGGGTCACTGCACCAAAATGGCCACCAGG
980963         143 GTGGAGACCAGCTGTGTGTCTGGGGTCACTGCACCAAAATGGCCACCAGG
W61032         124 GTGGAGACCAGCTGTGTGTCTGGGGTCACTGCACCAAAATGGCCACCAGG
853079          61 GTGGAGACCAGCTGTGTGTCTGGGGTCACTGCACCAAAATGGCCACCAGG
<DNA35814:     151 GTGGAGACCAGCTGTGTGTCTGGGGTCACTGCACCAAAATGGCCACCAGG

2054176        201 GGCAGCAATGGGACCATCTGTGACAACCAGAGGGACTGCCA
1756022        200 GG
980963         193 GGCAGCAATGGGACCATCTGTGACAACCAGAGGGACTGCCAGCCGGGGCT
W61032         174 GGCAGCAATGGGACCATCTGTGACAACCAGAGGGACTGCCAGCCGGGGCT
853079         111 GGCAGCAATGGGACCATCTGTGACAACCAGAGGGACTGCCAGCCGGGGCT
<DNA35814:     201 GGCAGCAATGGGACCATCTGTGACAACCAGAGGGACTGCCAGCCGGGGCT
```

FIGURE 3B

```
980963          243 GTGCTGTGCCTTCCAG
W61032          224 GTGCTGTGCCTTCCAGAGAGGCCTGCTGTTCCCTGTGTGCACACCCCTGC
853079          161 GTGCTGTGCCTTCCAGAGAGGCCTGCTGTTCCCTGTGTGCACACCCCTGC
<DNA35814:    251 GTGCTGTGCCTTCCAGAGAGGCCTGCTGTTCCCTGTGTGCACACCCCTGC

W61032          274 CCGTGGAGGN-GAGCTTTGCCATG
853079          211 CCGTGGAGGGCGAGCTTTGCCATG
<DNA35814:    301 CCGTGGAGGGCGAGCTTTGCCATG
```

FIGURE 4A

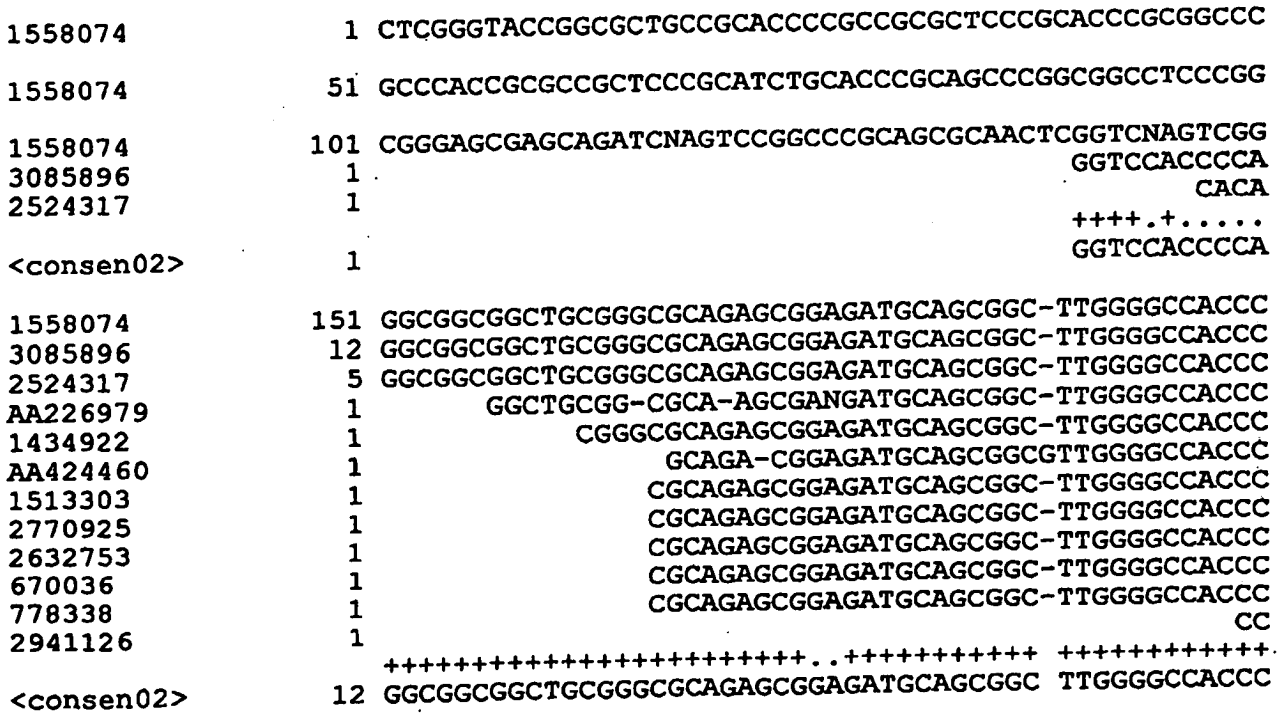
1558074 1 CTCGGGTACCGGCGCTGCCGCACCCCGCCGCGCTCCCGCACCCGCGGCCC
1558074 51 GCCCACCGCGCCGCTCCCGCATCTGCACCCGCAGCCCGGCGGCCTCCCGG
1558074 101 CGGGAGCGAGCAGATCNAGTCCGGCCCGCAGCGCAACTCGGTCNAGTCGG
3085896 1 GGTCCACCCCA
2524317 1 CACA
++++.+.....
<consen02> 1 GGTCCACCCCA
1558074 151 GGCGGCGGCTGCGGGCGCAGAGCGGAGATGCAGCGGC-TTGGGGCCACCC
3085896 12 GGCGGCGGCTGCGGGCGCAGAGCGGAGATGCAGCGGC-TTGGGGCCACCC
2524317 5 GGCGGCGGCTGCGGGCGCAGAGCGGAGATGCAGCGGC-TTGGGGCCACCC
AA226979 1 GGCTGCGG-CGCA-AGCGANGATGCAGCGGC-TTGGGGCCACCC
1434922 1 CGGGCGCAGAGCGGAGATGCAGCGGC-TTGGGGCCACCC
AA424460 1 GCAGA-CGGAGATGCAGCGGCGTTGGGGCCACCC
1513303 1 CGCAGAGCGGAGATGCAGCGGC-TTGGGGCCACCC
2770925 1 CGCAGAGCGGAGATGCAGCGGC-TTGGGGCCACCC
2632753 1 CGCAGAGCGGAGATGCAGCGGC-TTGGGGCCACCC
670036 1 CGCAGAGCGGAGATGCAGCGGC-TTGGGGCCACCC
778338 1 CGCAGAGCGGAGATGCAGCGGC-TTGGGGCCACCC
2941126 1 CC
++++++++++++++++++++++..++++++++++ ++++++++++
<consen02> 12 GGCGGCGGCTGCGGGCGCAGAGCGGAGATGCAGCGGC TTGGGGCCACCC

FIGURE 4B

```
1558074  200 TG-CTG
3085896   61 TG-CTGTGCCTGCTGCTGG-CGGCGGCGGTCCCCACGGNCCCCGCNCCCG
2524317   54 TG-CTGTGCCTGCTGCTGG-CGGCGGCGGTCCCCACGGCCCCCGCGCCCG
AA226979  42 TG-CTGTGCCTGCTGCTGG-CGGCGGCGGTCCCCACGGCCCCCGCGCCCG
1434922   39 TG-CTGTGCCTGCTGCTGG-CGGCGGCGGTCCCCACGGCCCCCGCGCCCG
AA424460  34 TGACTGTGCCTGCTGCTGG-CGGCGGCGGTCCCCACGGCCCCCGCGCCCG
1513303   35 TG-CTGTGCCTGCTGCTGG-CGGCGGCGGTCCCCACGGCCCCCGCGCCCG
2770925   35 TG-CTGTGCCTGCTGCTGG-CGGCGGCGGTCCCCACGGCCCCCGCGCCCG
2632753   35 TG-CTGTGCCTGCTGCTGG-CGGCGGCGGTCCCCACGGCCCCCGCGCCCG
670036    35 TG-CTGTGNCTGCTGCTGG-CGGCGGCGGTCCCCACGGNCCCCNCGNCCG
778338    35 TG-CTGTGCCTGCTGCTGG-CGGCGGCGGTCCCCACGGCCCCCGCGCCCG
2941126    3 TG-CTGTGCCTGCTGCTGG-CGGCGGCGGTCCCCACGGCCCCCGCGCCCG
2818603    1
                ++ +++++.++++++++++ ++++++++++++++++.++++.+..+++
<consen02>  61 TG CTGTGCCTGCTGCTGG CGGCGGCGGTCCCCACGGCCCCCGCGCCCG

3085896   109 CTCCGANGGCGACCNNGGCTCCAGTCAAGCCCGGCCCGGCTCTCAGCTAC
2524317   102 CTCCGACGGCGACCTCGGCTCCAGTCAAGCNCGGCCCGGCT
AA226979   90 CTCCGACGGCGACCTCGGCTCCAGTCAAGCCCGGCCCGGCTCTCAGCTAC
1434922    87 CTCCGACGGCGACCTCGGCTCCAGTCAAGCCCGGCCCGGCTCTCAGCTAC
AA424460   83 CTCCGACGGCGACCTCGGCTCCAGTCAAGCCCGGCCCGGCTCTCAGCTAC
1513303    83 CTCCGACGGCGACCTCGGCTCCAGTCAAGCCCGGCCCGGCTCTCAGCTAC
2770925    83 CTCCGACGGCGACCTCGGCTCCAGTCAAGCCCGGCCCGGCTCTCAGCTAC
2632753    83 CTCCGACGGCGACCTCGGCTCCAGTCAAGCCCGTNCCGGGTCTNAGNTAC
670036     84 NTCCGACGNCGACCTTGGCTCCAGTCAAGCCCGGCCCGGNTNTNAGCTAC
778338     83 NTCCGACGGCGACCTCGGCTCCAGTCAAGCCCGGCCCGGCTCTCAGCTAC
2941126    51 CTCCGACGGCGACCTCGGCTCCAGTCAAGCCCGGCCCGGCTCTCAGCTAC
2818603    18 CTCCGACGGCGACCTCGGCTCCAGTCAAGCCCGGCCCGGCTCTCAGCTAC
                .++++++.+..++++++..++++++++++++++.++..++++.+.+..++.+++
<consen02> 109 CTCCGACGGCGACCTCGGCTCCAGTCAAGCCCGGCCCGGCTCTCAGCTAC
```

FIGURE 4C

3085896	159	C-CGCAGGAG-GAGGCCACCCTCAATGAGATGTTCCGCGAAGTTGAGGAA
AA226979	140	CGCGCAGGAG-GAGGCCACCCTCAATGAGATGTTCCGCGAGGTTGAGGAA
1434922	137	C-CGCAGGAG-GAGGCCACCCTCAATGAGATGTTCCGCGAG-TTGAGGAA
AA424460	133	C-CGCAGGAGCGAGGCCACCCTCAATGAGATGTTCCGCGAGGTTGAGGAA
1513303	133	C-CGCAGGAG-GAGGCCACCCTCAATGAGATGTTCCGCGAGGTTGAGNGA
2770925	133	C-CGCAGGAG-GAGGCCACCCTCAATGAGATGTTCCGCGAGGTTGAGGAA
2632753	133	C-CGCAG
670036	134	C-CGNAGGAG-GAGGCCACCNTNANTGAGATGTTTCN-GAGGTTGAGGAA
778338	133	C-CGCAGGAG-GAGGCCACCCTCAATGAGATGTTCCGCGAGGTTGAGGAA
2941126	101	C-CGCAGGAG-GAGGCCACCCTCAATGAGATGTTCCGCGAGGTTGAGGAA
2818603	68	C-CGCAGGAG-GAGGCCACCCTCAATGAGATGTTCCGCGAG-TTGAGGAA
		++++.++++++++++++++.+.+.+++++++++.+.+++.++++++..+
<consen02>	159	CGCGCAGGAGCGAGGCCACCCTCAATGAGATGTTCCGCGAGGTTGAGGAA
3085896	207	CTGATGGAGG--ACACGCAGCACAAATTGCGCACGC-GGT-GG-AAGAGA
AA226979	189	CTGATGGAGG--ACACGCAGCACAAATTG-GCAC-C-GGT-GG-AAGAGA
1434922	184	CTGATGGAGG--ACACGCAGCACAAATTGCGCAAGCCGGT-GG-AAGAGA
AA424460	182	CTGATGGAGG--ACACGCAGCACAAATTGCGCA-GC-GGT-GGGAAGAGA
1513303	181	ATGATGNAGGGAAAACGNAGCANAAATT
2770925	181	-TGATGGAGG--ACACGCAGCACAAATTGCGCACGC-GGT-GG-AAGAGA
670036	181	CTTATGGAGG--ACACGTAGNACAAATTNNGCAGGC-GGTTGG-AAGAGA
778338	181	CTGATGGAGG--ACACGNAGNACAAATTGCGCAGGC-GGT-GG-AAGAGA
2941126	149	CTGATGGAGG--ACACGCAGCACAAATTGCGCANGC-GGT-GG-AAGAGA
2818603	115	CTGATGGAGG--ACACGCAGCACAAATTGCGCACGC-GGT-GG-AAGAGA
		.+.+++.+++ +.+++.++.+.+++++..+++.++ +++ ++ ++++++
<consen02>	209	CTGATGGAGG ACACGCAGCACAAATTGCGCACGC GGT GG AAGAGA

FIGURE 4D

```
3085896          252 T-GGAGGCAGAAGAAGCTG-TGCTAAANC
AA226979         232 T-GGAGGCAGAAGAAGCTGCTGCTAAAGCATCATCAGAAGTGAACCTGGC
1434922          230 T-GAAGCAAAAAAACTT
AA424460         227 T-GGAGGCAGAAGAAGCTGCTG
2770925          225 T-GGAGGCAGAAGAAGCTGCTGCTAAAGCA
670036           227 TTGGAGTCAGAAGANGNTGNTTGTAAAGNATCATNAGT-GTGAACCTTGN
778338           226 T-GGAGGCAGAAGAAGCTGCTAA
2941126          194 T-GGAGGCAGAAGAAGCTGCTGCTAAAGCATCATCAGAAGTGAACCTGGC
2818603          160 T-GGAGGCAGAAGAAG-TGCTGCTAAAGCATCATCAG
1452718            1      GGCAGAAGAAGCTGCTGCTAAAGCATCATCAGAAGTGAACCTGGC
2845278            1           GAAGAAGCTGCTGCTAAAGCATCATCAGAAGTGAACCTGGC
                     + +.++..+.++.+...++.+..++++..+++++.++.+++++++++.+.
<consen02>       254 T GGAGGCAGAAGAAGCTGCTGCTAAAGCATCATCAGAAGTGAACCTGGC

AA226979         281 AAACTTACCTCCCAGCTATCACAATGAGACCAACACAGACACGAAGGTTG
670036           276 AAANTTANCTTC
2941126          243 AAACTTACCTCCCAGCTATCACAA
1452718           46 AAACTTACCTCCCAGCTATCACAATGAGACCAACACAGACACGAAGGTTG
2845278           42 AAACTTACCTCCCAGCTATCACAATGAGACCAACACAGACACGAAGGTTG
923337             1               AGCTATCACAATGAGACCAACACAGACACGAAGGTTG
T08793              1                                                TTG
                     +++.+++.++.++++++++++++++++++++++++++++++++++++++++
<consen02>       303 AAACTTACCTCCCAGCTATCACAATGAGACCAACACAGACACGAAGGTTG
```

FIGURE 4E

```
AA226979          331 GAAATAATACCATCCATGTGCACCGAGAA
1452718            96 GAAATAATACCATCCATGTGCACCGAGAAATTCACAAGATAACCAACAAC
2845278            92 GAAATAATACCATCCATGTGCACCGAGAAATTCACAAGATAACCAACAAC
923337             38 GAAATAATACCATCCATGTGCACCGAGAAATTCACAAGATAACCAACAAC
T08793              4 GAAATAATACCATCCATGTGCACCGAGAAATTCACAAGATAACCAACAAC
T30923              1         TACCATCCATGTGCACCGAGAAATTCACAAGATAACCAACAAC
981001              1             ATCCATGTGCACCGAGAAATTCACAAGNTAACCAACAAC
261347              1                     CCAGAGGAGAAATTCACAAGATAACCAACAAC
T31076              1                           GAGAAATTCACAAGATAACCAACAAC
                      ++++++++++++++++++.....++++++++++++.++++++++++
<consen02>        353 GAAATAATACCATCCATGTGCACCGAGAAATTCACAAGATAACCAACAAC

1452718           146 CAGACTGGACAAATGGTCTTTTCAGAGACAGTTATCACATCTGTGGGAGA
2845278           142 CAGACTGGACAAATGGTCTTTTCAGAGACAGTTATCACATCTGTNGGAGA
923337             88 CAGACTGGACAAATGGTCTTTTCAGAGACAGTTATCACATCTGTGGGAGA
T08793             54 CAGACTGGACAAATGGTCTTTTCAGAGACAGTTATCACATCTGTGGGAGA
T30923             44 CAGACTGGACAAATGGTCTTTTCAGAGACAGTTATCACATCTNTGGGAGA
981001             40 CAGACTNGACAAATGGTCTTTTCAGAGACAGTTATNACANCTGTGGGAGA
261347             33 CAGACTGGACAAATGGTCTTTTCAGAGACAGTTATCACATCTNTGGGAGA
T31076             27 CAGACTGGACAAATGGTCTTTTCAGAGACAGTTATCACATCTGTGGGAGA
1756022             1                                   ATCACATNTGTGGGAGA
1549807             1                                     CACATCTGTNGGANA
DNA35814.init       1                                          CTGTGGGAGA
2054176             1                                          CTGTGGGAGA
980963              1                                                  GA
                      ++++++.++++++++++++++++++++++++++.+++..+.+.+++.+
<consen02>        403 CAGACTGGACAAATGGTCTTTTCAGAGACAGTTATCACATCTGTGGGAGA

1452718           196 CGAAGAAGGCAGAAGGAGCCACGAGTGCATCATCGACGAGGACTGT-GGG
2845278           192 CGAAGAAGGCAGAAGGAGCCACGAGTGCATCATCGACGAGGACTGT-GGG
923337            138 CGAAGAAGGCAGAAGGAGCCACGAGTGCATCATCGACGAGGACTGT-GGG
T08793            104 CGAAGAAGGCAGAAGGNGCCACGAGTGCATCATCGACGAGGACTGT-GGG
T30923             94 CGAAGAAGGCAGAAGGAGCCACGAGTGCATCATCGACGAGGACTNT-GGG
981001             90 CGGAGAAGGCAGAAGGAGNCACGGGTNCAATATTGGCGGGGGCTTTTGGG
```

FIGURE 4F

```
261347              83 CGAAGAAGGCAGAAGGAGCCACGAGTGCATNATCGACGAGGACTGT-NGG
T31076              77 CGAAGAAGGCAGAAGGAGCCACGAGTGCATCATCGACGAGGACTNT-GGG
1756022             18 CGAAGAAGGCAGAAGGAGCCACGAGTGCATCATCGACGAGGACTGT-GGG
1549807             16 CGAANAAGGCANAAGGAGCCACGAGTGCATCATCGACGAGGACTGT-NGG
DNA35814.init       11 CGAAGAAGGCAGAAGGAGCCACGAGTGCATCATCGACGAGGACTGT-GGG
2054176             11 CGAAGAAGGCAGAAGGAGCCACGAGTGCATCATCGACGAGGACTGT-GGG
980963               3 CGAAGAAGGCAGAAGGAGCCACGAGTGCATCATCGACGAGGACTGT-GGG
2901484              1             GAAGGAGCCACGAGTGCATCATCGACGAGGACTGT-GGG
W61032               1                       GCCACGAGTGCATCATCGACGAGGACTGT-GGG
                       ++.+.++++++.++++.+.++++.++.++..++.+.++.++.++.++.+ .++
<consen02>         453 CGAAGAAGGCAGAAGGAGCCACGAGTGCATCATCGACGAGGACTGT GGG

1452718            245 CCCAG
2845278            241 CCCAGCATGTACTGCCAGT
923337             187 CCCAGCATGTACTGCCAGTTTGCCAGCTTCCAGTACACCTGCCAGCCATG
T08793             153 CCCAGCATGTACTGCCAGTTTGCCAGCTTCCAGTACACCTGCCAGCCATG
T30923             143 CCCAGCATGTACTGCCAGTTTGCCAGCTTCCAGTACACCTGCCAGCCATG
981001             140 CCCAGNATNTACTTCCAGTTTTCCAG-TTCCAGTANAC-TNCCAGCCATG
261347             132 CCCAGCATGTACTGCCAGTTTGCCAGCTTCCAGTACACCTNCCAGCCATG
T31076             126 CCCAGCATGTACTGCCAGTTTNCCAGCTTCCAGTACACCTGCCAGCCATG
1756022             67 CCCAGCATGTACTGCCAGTTTGCCAGCTTCCAGTACACCTGCCAGCCATG
1549807             65 CCCAGCATGTACTGCCAGTTTGCCAGCTTCCAGTACACCTGCCAGCCATG
DNA35814.init       60 CCCAGCATGTACTGCCAGTTTGCCAGCTTCCAGTACACCTGCCAGCCATG
2054176             60 CCCAGCATGTACTGCCAGTTTGCCAGCTTCCAGTACACCTGCCAGCCATG
980963              52 CCCAGCATGTACTGCCAGTTTGCCAGCTTCCAGTACACCTGCCAGCCATG
2901484             39 CCCAGCATGTACTGCCAGTTTGCCAGCTTCCAGTACACCTGCCAGCCATG
W61032              33 CCCAGCATGTACTGCCAGTTTGCCAGCTTCCAGTACACCTGCCAGCCATG
2595269              1  CCAGCATGTACTGCCAGTTTGCCAGCTTCCAGTACACCTGCCAGCCATG
853079               1                                    AGTACACCTGCCAGCCATG
863643               1                                              CAANCATG
                       +++++.++.++++.+++++++.++++++++++++.++++.+++..++++
<consen02>         502 CCCAGCATGTACTGCCAGTTTGCCAGCTTCCAGTACACCTGCCAGCCATG
```

60863435.102992

FIGURE 4G

```
923337         237 CCGGGGCCAGAGGATGCTCTGCACCCGGGACAGTGAGTGCTGTGGAGACC
T08793         203 TNGGGGCCAGAGGATGCTCTGCACCCGGGACAGT
T30923         193 CCGGGGCCAGAGGATGCTCTNCACCCGGGACAGTGAGTGCTGTGGAGACC
981001         188 CCGGGGCCANAGGAT-CTTTTAACC-GGG-CANTGAGTNCTNTG-AGACC
261347         182 NCGGGGCCAGAGGATGCTCTNCACCCGGGACAGTTAGTTCTGTNGAGACC
T31076         176 CCGGGGCCAGAGGATGCTCTGCACCCGGGACAGTGAGTGCTGTGGAGACC
1756022        117 CCGGGGCCAGAGGATGCTCTGCACCCGGGACAGTGAGTGCTGTGGAGACC
1549807        115 CCGGGGCCAGAGGATGCTCTGCACCCGGGACAGTNAGTGCTGTGGAGACC
DNA35814.init  110 CCGGGGCCAGAGGATGCTCTGCACCCGGGACAGTGAGTGCTGTGGAGACC
2054176        110 CCGGGGCCAGAGGATGCTCTGCACCCGGGACAGTNAGTGCTGTGGAGACC
980963         102 CCGGGGCCAGAGGATGCTCTGCACCCGGGACAGTGAGTGCTGTGGAGACC
2901484         89 CCGGGGCCAGAGGATGCTCTGCACCCGGGACAGTGAGTGCTGTGGAGACC
W61032          83 CCGGGGCCAGAGGATGCTCTGCACCCGGGACAGTGAGTGCTGTGGAGACC
2595269         50 CCGGGGCCAGAGGATGCTCTGCACCCGGGACAGTGAGTGCTGTGGAGACC
853079          20 CCGGGGCCAGAGGATGCTCTGCACCCGGGACAGTGAGTGCTGTGGAGACC
863643           9 ACGGGGNCAGAGGATGCTCTGCANCCGGGACAGTGAGTGCTGTGGAGANC
1579793          1                           ACCCGGGACAGTGAGTGCTGTGGAGACC
2482612          1                                                  AGACC
R14945           1                                                     CC
                   ..++++.++.++++++++.+..+.++++++++.+.+++.++.+.++++.+
<consen02>     552 CCGGGGCCAGAGGATGCTCTGCACCCGGGACAGTGAGTGCTGTGGAGACC

923337         287 AGCTGTGTGTCTGGGGTCACTGCACCAAAATGG
T30923         243 AGCTGTGTGTCTGGGGTCACT
981001         234 AG-TGTGTGTTTGGG-T-ANTNCACCAAA-TGGC-ACCAGGG-CAG-AAT
261347         232 AGCTNTNTGTCTGGGGTCACTTCACCAAAATNGCCACCAGGGGCAGCAAT
T31076         226 AGCTGTGTGTCTG
1756022        167 AGCTGTGTGTCTGGGGTCACTGCACCAAAATGGCCACCAGGGG
1549807        165 AGCTGTGTGTCTGGGGTCACTGCACCAAAATGGCCACCAGG
DNA35814.init  160 AGCTGTGTGTCTGGGGTCACTGCACCAAAATGGCCACCAGGGGCAGCAAT
2054176        160 AGCTGTGTGTCTGGGGTCACTGCACCAAAATGGCCACCAGGGGCAGCAAT
980963         152 AGCTGTGTGTCTGGGGTCACTGCACCAAAATGGCCACCAGGGGCAGCAAT
2901484        139 AGCTGTGTGTCTGGGGTCACTGCACCAAAATGGCCACCAGGGGCAGCAAT
W61032         133 AGCTGTGTGTCTGGGGTCACTGCACCAAAATGGCCACCAGGGGCAGCAAT
```

FIGURE 4H

```
2595269        100 AGCTGTGTGTCTGGGGTCA-TGNACCAAAATGGCCACCAGGGGCAGCAAT
853079          70 AGCTGTGTGTCTGGGGTCACTGCACCAAAATGGCCACCAGGGGCAGCAAT
863643          59 AGCTGTGTGTCTGGGGTCACTGNACCAAAATGGACACCAGGGGCAGCAAT
1579793         29 AGCTGTGTGTCTGGGGTCACTGCACCAAAATGGCCACCAGGGGCAGCAAT
2482612          6 AGCTGTGTGTCTGGGGTCACTGCACCAAAATGGCCACCAGGGGCAGCAAT
R14945           3 AGCTGTGTGTCTGGGGTCACTGCACCAAAATGGCCACCAGGGGCAGCAAT
1349102          1                   GGTCACTGCACCAAAATGGCCACCAGGGGCAGCAAT
2656314          1                         CTGCACCAAAATGGCCACCAGGGGCAGCAAT
2658126          1                         CTGCACCAAAATGGCCACCAGGGGCAGCAAT
1478240          1                         CTGCACCAAAATGGCCACCAGGGGCAGCAAT
2524907          1                                        GGCCACCAGGGGCAGCAAT
                   ++++.+.+++.++++++++.+..++++++++.+.+++++++++++++++
<consen02>     602 AGCTGTGTGTCTGGGGTCACTGCACCAAAATGGCCACCAGGGGCAGCAAT

981001         277 TGGACCAT-TGTGACAAC-AGAGGN-CTNCCAGC--GGGCTNTG-T-GTG
261347         282 NGGNCCATCTNTNANAACCAGAGGGNCTNC-AGNCGGGGCTGTGCT-GTN
DNA35814.init  210 GGGACCATCTGTGACAACCAGAGGGACTGCCAGCCGGGGCTGTGCT-GTG
2054176        210 GGGACCATCTGTGACAACCAGAGGGACTGCCA
980963         202 GGGACCATCTGTGACAACCAGAGGGACTGCCAGCCGGGGCTGTGCT-GTG
2901484        189 GGGACCATCTGTGACAACCAGAGGGACTGCCAGCCGGGGCTGTGCT-GTG
W61032         183 GGGACCATCTGTGACAACCAGAGGGACTGCCAGCCGGGGCTGTGCT-GTG
2595269        149 GGGACCATCTGTGACANCCAGAGGGANTGNNANNCGGGGNTTNTGT
853079         120 GGGACCATCTGTGACAACCAGAGGGACTGCCAGCCGGGGCTGTGCT-GTG
863643         109 GGGACCATCTGTGACAACCAGAGGGACTGNCANNCGGGGCTGTGCT-GTG
1579793         79 GGGACCATCTGTGACAACCAGAGGGACTGCCAGCCGGGGCTGTGCT-GTG
2482612         56 GGGACCATCTGTGACAACCAGAGGGACTGCCAGCCGGGGCTGTGCT-GTG
R14945          53 GGGACCATCTGTGACAACCAGAGGGACTGCCAGCCGGGGCTGTGCT-GTG
1349102         37 GGGACCATCTGTGACAACCAGAGGGACTGCCAGCCGGGGCTGTGCT-GTG
2656314         32 GGGACCATCTGTGACAACCAGAGGGACTGCCAGCCGGGGCTGTGCT-GTG
2658126         32 GGGACCATCTGTGACAACCAGAGGGACTGCCAGCCGGGGCTGTGCT-GTG
1478240         32 GGGACCATCTGTGACAACCAGAGGGACTGCCAGCCGGGGCTGTGCT-GTG
2524907         20 GGGACCATCTGTGACAACCAGAGGGACTGCCAGCNGGGGCTGTGCT-GTG
925806           1       ATCTNTGACAACCAGAGGGACTNCCAGCCGGGGCTNTGCT-NTG
W45085           1                  ACCAGAGGGACTGCCAGCCGGGG-TGTGCTCGTG
```

FIGURE 4I

```
1865280              1                                                CGGGGCTGTGCT-GTG
                                                                      .++.++++++.+.+.+.+++++++...+...+...++++.+....+ .+.
<consen02>         652 GGGACCATCTGTGACAACCAGAGGGACTGCCAGCCGGGGCTGTGCT GTG

981001             320 CTTTC-AGAGAGGCTG
261347             330 CCTTCCNGAGAGGCCTGCTGTTCCCTGTGTNNAAACCCTT-CNCNTNGAG
DNA35814.init      259 CCTTCCAGAGAGGCCTGCTGTTCCCTGTGTGCACACCCCTGCCCGTGGAG
980963             251 CCTTCCAG
2901484            238 CCTTCCAGAGAGGCCTG
W61032             232 CCTTCCAGAGAGGCCTGCTGTTCCCTGTGTGCACACCCCTGCCCGTGGAG
853079             169 CCTTCCAGAGAGGCCTGCTGTTCCCTGTGTGCACACCCCTGCCCGTGGAG
863643             158 NCTTACAGAGAGGACTTCTNTTG
1579793            128 CCTTCCAGAGAGGCCTGCTGTTCCCTGTGTGCACACCCCTGCCCGTG
2482612            105 CCTTCCAGAGAGGCCTGCTGTTCCCTGTGTGCACACCCCTGCCCGTGGAG
R14945             102 CCTTCCAGAGAGGCCTGCTGTTCCCTGTGTGCACACCCCTGCCCGTGGAG
1349102             86 CCTTCCAGAGAGGCCTGCTGTTCCCTGTGTGCACACCCCTGCCCGTGGAG
2656314             81 CCTTCCAGAGAGGCCTGCTGTTCCCTGTGTGCACACCCCTGCCCGTGGAG
2658126             81 CCTTCCAGAGAGGCCTGCTGTTCCCTGTGTGCACACCCCTGNCCGTGGAG
1478240             81 CCTTCCAGAGAGGCCTGCTGTTCCCTGTGTGCACACCCCTGCCCGTGGAG
2524907             69 CCTTCCAGAGAGGCCTGCTGTTCCCTGTGTGCACACCCCTGCCCGTGGAG
925806              44 CCTTCCAGAGAGGCCTNCTGTTCCCTGTGTGCACACCCNTGCCCGTGGAG
W45085              34 CCTTCCAGAGAGGCCTGCTGTTCCCTGTGTGCACACCCCTGCCCGTGGAG
1865280             16 CCTTCCAGAGAGGCCTGCTGTTCCCTGTGTGCACACCCCTGCCCGTGGAG
                       ..++.+.++++++....++.++.+++++++..+.++++.++..+.+.+++
<consen02>         701 CCTTCCAGAGAGGCCTGCTGTTCCCTGTGTGCACACCCCTGCCCGTGGAG

261347             379 GGN-A-GCTTTNCATT
DNA35814.init      309 GGCGA-GCTTTGCCATG
W61032             282 GN-GA-GCTTTGCCATGACCCCGCCAGCCGGCTTCTGGACCTCATCACCT
853079             219 GGCGA-GCTTTGCCATG
2482612            155 GGCGA-GCTTTGCCATGACCCCGCCAGCCGGCTTCTGGACCTCATCACCT
R14945             152 GGAN--GCTTTGCCATGACCCCGCCAGCCGGCTTCTGGACCTCATCACCT
1349102            136 GGCGA-GCTTTGCCATGACCCCGCCAGCCGGCTTCTGGACCTCATCACCT
2656314            131 GGCGA-GCTTTGCCATGACCCCGCCAGCCGGCTTCTGGACCTCATCACCT
```

FIGURE 4J

```
2658126          131 GGCGA-GCTTTGCCATGACCCCGCCAGCCGGCTTCTGGACCTCATCACCT
1478240          131 GGCGA-GCTTTGCCATGACCCCGCCAGCCGGCTTCTGGACCTCATCACCT
2524907          119 GGCGA-GCTTTGCCATGACCCCGCCAGCNGGCTTCTGGACCTCATCACCT
925806            94 GGCGA-GNTTTGCCATGACCCCGNCAGCCGGNTTNTNGACCTCATNACCN
W45085            84 CG-GACGCTTTGC-ATGACCCCGCCAGCCGGCTTCTGGACCTCATCACCT
1865280           66 GGCGA-GCTTTGCCATGACCCCGCCAGCCGGCTTCTGGACCTCATCACCT
                     ....+ +.+++.+..++++++++.++++.++.++.+.++++++++.+++.
<consen02>       751 GGCGA GCTTTGCCATGACCCCGCCAGCCGGCTTCTGGACCTCATCACCT

W61032           330 GGG-AGCTAGAGCCTGATGGAGCCTTGG-ACCGATGCCCTTGTGCCAGTG
2482612          204 GGG-AGCTA
R14945           200 GGGGAGCTAGAGCCTGATGGAGCCTTGGGACCGATGCCCTTGTGCCAGTG
1349102          185 GGG-AGCTAGAGCCTGATGGAGCCTTGG-ACCGATGCCCTTGTGCCAGTG
2656314          180 GGG-AGCTAGAGCCTGATGGAGCCTTGG-ACCGATGCCCTTGTGCCAGTG
2658126          180 GGG-AGCTAGAGCCTGATGGAGCCTTGG-ACCGATGCCCTTGTGCCAGTG
1478240          180 GGG-AGCTAGAGCCTGATGGAGCCTTGG-ACCGATGCCCTTGTGCCAGTG
2524907          168 GGG-AGCTAGAGCCTGATGGAGCCTTGG-ACCGATGCCCTTGTGCCAGTG
925806           143 NGG-AGCTAGAGCCTGATGGAGCCTTGG-ACCGANGCCNTTGTGCCAGTG
W45085           132 GGG-AGCTAGAGCCTGATGGAGCCTTGG-ACCGATGCCCTTGTGCCAGTG
1865280          115 GGG-AGCTAGAGCCTGATGGAGCCTTGG-ACCGATGCCCTTGTGCCAGTG
865729             1                                CGATGCCCTTGTGCCAGTG
2760910            1                                     CCCTTGTGCCAGTA
2496946            1                                        TTGTGCCAGTN
                     .++ ++++++++++++++++++++++ +++++.+++.++++++++++.
<consen02>       800 GGG AGCTAGAGCCTGATGGAGCCTTGG ACCGATGCCCTTGTGCCAGTG

W61032           378 GCCTCCTCT-GCCAGCCCCACAGCCACAGCCTGG-TGTATGT-GTGCAA-
R14945           250 GCCTCCTCTTGCCAGCCCCACAGCCACAGCCTGGGTGTATGTTGTGCAAA
1349102          233 GCCTCCTCT-GCCAGCCCCACAGCCACAC
2656314          228 GCCT
2658126          228 GCCTCCTCT-GCCAGCCC
1478240          228 GCCTCCT
2524907          216 GCCTCCTCT-GCCAGCNCCACA
925806           191 GCCTACTAN-GNCAGCCCCACAGCCACAGCCTGG-TNTATGT-GTGCAA-
```

FIGURE 4K

```
W45085          180 GC-TCCTCT-GCCAGCCCCACAGCCACAGCCTGG-TGTATGT-GTGCAA-
1865280         163 GCCTCCTCT-GCCAGCCCCACAGCCACAGCCTGG-TGTATGT-GTGCAA-
865729           20 GNCTNCTNT-GCCAGCCCCACAGCCACAGCCTGG-TGTATGT-GTGCAA-
2760910          15 GCCTCCTCT-GCCAGCCCCACAGCCACAGCCTGG-TGTATGT-GTGCAA-
2496946          12 GCCTCCTCT-GCCAGCCCCACAGCCACAGCCTGG-TGTATGT-GTGCAA-
2223827           1            GCCAGCCCCACAGCCACAGCCTGG-TGTATGT-GTGCAA-
2226521           1            GCCAGCCCCACAGCCACAGCCTGG-TGTATGT-GTGCAA-
1466083           1                  CCCACAGCCACAGCCTGG-TGTATGT-GTGCAA-
927569            1                  CCCACAGCCACAGCCTGG-TGTATGT-GTGCAA-
                    +.++.++.. +.++++.++++++++++.+++++ +.+++++ ++++++
<consen02>      848 GCCTCCTCT GCCAGCCCCACAGCCACAGCCTGG TGTATGT GTGCAA

W61032          424 GCCGACCTTCGT-GGGGAGCCGTGACCAAGATGGGGAGATC-TGCTGCCC
R14945          300 GCCGACCTTCGTNGGGGAACCGTGACCAAGATGGGGGAGATTCTT
925806          237 GCCGACCTTAGT-GGGGAGCCGTGACCAAGATGGGGAGATN-TG-TGCCC
W45085          225 GCCGACCTTCGT-GGGGAGCCGTGACCAAGATGGGGAGATCCTGCTGCCC
1865280         209 GCCGACCTTCGT-GGGGAGCCGTGACCAAGATGGG
865729           66 GCCGACCTTCGT-GGGGAGCCGTGACCAAGATGGGGAGATCCTGCTGCCC
2760910          61 GCCGACCTTCGT-GGGGAGCCGTGACCAAGATGGGGAGATCCTGCTGCCC
2496946          58 GCCGACCTTCGT-GGGGAGCCGTGACCAAGATGGGGAGATCCTGCTGCCC
2223827          38 GCCGACCTTCGT-GGGGAGCCGTGACCAAGATGGGGAGATCCTGCTGCCC
2226521          38 GCCGACCTTCGT-GGGGAGCCGTGACCAAGATGGGGAGATCCTGCTGCCC
1466083          32 GCCGACCTTCGT-GGGGAGCCGTGACCAAGATGGGGAGATCCTGCTGCCC
927569           32 GCCGACCTTCGT-GGGGAGCCGTGACCAAGATGGGGAGATCCTGCTGCCC
354028            1                     CCGTGACCAAGATGGGGAGATCCTGCTGCCC
2277563           1                     CCGTGACCAAGATGGGGAGATCCTGCTGCCC
1289239           1                      CGTGACCAAGATGGGGAGATCCTGCTGCCC
AA031480.RC       1                                        TCCTGCTGCCC
                    +++++++++.++ +++++.++++++++++++++++.........+++++
<consen02>      894 GCCGACCTTCGT GGGGAGCCGTGACCAAGATGGGGAGATCCTGCTGCCC

W61032          472 A-GAGAGGTCCC-GAT-GAGTATGAAGTTGGAA--CTTCATGGAGGAGGT
925806          284 A-GAGAGGTTCC-GAT-NAGTATGA
W45085          274 AANAAAGGTCCCCGATTGAGTATGAAGTTGGCAAGCTTCATGGAAGGAAN
```

FIGURE 4L

```
865729          115 A-GAGAGGTCCCCGAT-GAGTATGAAGTTGGCA-GCTTCATGGAGGAGGT
2760910         110 A-GAGAGGTCCCCGAT-GAGTATGAAGTTGGCA-GCTTCATGGAGGAGGT
2496946         107 A-GAGAGGTCCCCGAT-GAGTATGAAGTTGGCA-GCTTCATGGAGGAGGT
2223827          87 A-GAGAGGTCCCCGAT-GAGTATGAAGTTGGCA-GCTTCATGGAGGAGGT
2226521          87 A-GAGAGGTCCCCGAT-GAGTATGAAGTTGGCA-GCTTCATGGAGGAGGT
1466083          81 A-GAGAGGTCCCCGAT-GAGTATGAAGTTGGCA-GCTTCATGGAGGAGGT
927569           81 A-GAGAGGTCCCCGAT-GAGTATGAAGTTGGCA-GCTTCATGGAGGAGGT
354028           32 A-GAGAGGTCCCCGAT-GAGTATGAAGTTGGCA-GCTTCATGGAGGAGGT
2277563          32 A-GAGAGGTCCCCGAT-GAGTATGAAGTTGGCA-GCTTCATGGAGGAGGT
1289239          31 A-GAGAGGTCCCCGAT-GAGTATGAAGTTGGCA-GCTTCATGGAGGAGGT
AA031480.RC      12 A-GAGAGGTCCCCGAT-GAGTATGAAGTTGGCA-GCTTCATGGAGGAGGT
R52311            1                                 NAAGTTGGCA-GCTTCATGGAGGAGGT
T33818            1                                            CTTCATGGAGGAGGT
                    + .+.++++.++++++ .+++++.+++++++.+ ++++++++++.+....
<consen02>      943 A GAGAGGTCCCCGAT GAGTATGAAGTTGGCA GCTTCATGGAGGAGGT

W61032          517 NCGCAAGAA-CTTGAAGACTTG-AGAGGAGCTT-ACTGAA-NAAT
W45085          324 GG
865729          162 GCGCCAGGAGCTNGAGGACCTGGAGAGGAGCCTGACTNAA-GAGATGGCG
2760910         157 GCGCCAGGAGCTGGAGGACCTGGAGAGGAGCCTGACTGAA-GAGATGGCG
2496946         154 GCGCCAGGAGCTGGANGACCTGGAGAGGAGCCTGACTGAA-GAGATGGCG
2223827         134 GCGCCAGGAGCTGGAGGACCTGGAGAGGAGCCTGACTGAA-GAGATGGCG
2226521         134 GCGCCAGGAGCTGGAGGACCTGGAGAGGAGCCTGACTGAA-GAGATGGCG
1466083         128 GCGCCAGGAGCTGGAGGACCTGGAGAGGAGCCTGACTGAA-GAGATGGCG
927569          128 GCGCCAGGAGCTGGAGGACCTGGAGAGGAGCCTGACTGAA-GAGATGGCG
354028           79 GCGCCAGGAGCTGGAGGACCTGGAGAGGAGCCTGACTGAA-GAGATGGCG
2277563          79 GCGCCAGGAGCTGGAGGACCTGGAGAGGAGCCTGACTGAA-GAGATGGCG
1289239          78 GCGCCAGGAGCTGGAGGACCTGGAGAGGAGCCTGACTGAA-GAGATGGCG
AA031480.RC      59 GCGCCAGGAGCTGGAGGACCTGGAGAGGAGCCTGACTGAA-GAGATGGCG
R52311           27 GCGCCAGGAGCTGGAGGACCTGGAGAGGAGCCTGACTGAA-GAGATGGCG
T33818           16 GCGCCAGGAGCTGGAGGACCTGGAGAGGAGCCTGACTGAA-GAGATGGCG
3087465           1 GCGCCAGGAGCTGGAGGACCTGGAGAGGAGCCTGACTGAA-GAGATGGCG
2525451           1           CTGGAGGACCNGGAGAGGAGCCTGACTGAA-GANATGGCG
C05296            1                     CTGGNAGAAGGAGCCTGACTGAAAGAGATGGCG
```

FIGURE 4M

```
C05172                1                                  CTGGNGAGGAGCCTGACTGAA-GAGATGGCG
2189372               1                                   TGGAGAGGAGCCTGACTGAA-GAGATGGCG
HSC0CH081             1                                       GAGGAGCCTGACTGAA-GAGATGGCG
1369095               1                                                              CG
                        ..++.++.++++.++.+........++++++.+++++.++ .+..+++++
<consen02>          990 GCGCCAGGAGCTGGAGGACCTGGAGAGGAGCCTGACTGAA GAGATGGCG

865729              211 CTNAGGGAGCCTTCGGCTNNCGACGGTGCACTNNTNGGAGGGGAAGA
2760910             206 CTGGGGGAGCCTGCGGCTGCCGCCGCTGCACTGCTGGGAGGGGAAGAGAT
2496946             203 CTGGGGGAGCCTGCGGCTGCCGCCGCTGCACTGCTGGGAAGGGAAGAGAT
2223827             183 CTGAGGGAGCCTGCGGCTGCCGCCGCTGCACTGCTGGGAGGGGAAGAGAT
2226521             183 CTGAGGGAGCCTGCGGCTGCCGCCGCTGCACTGCTGGGAGGGGAAGAGAT
1466083             177 C
927569              177 CTGGGGGAGCCTGCGGCTGCCGCCGCTGCACTGCTGGGAGGGGAAGAGAT
354028              128 CTGAGGGAGCCTGCGGCTGCCGCCGCTGCACTGCTGGGAGGGGAAGAGAT
2277563             128 CTGGGGGAGCCTGCGGCTGCCGCCGCTGCACTGCTGGGAGGGGAAGAGAT
1289239             127 CTGGGGGAGCCTGCGGCTGCCGCCGCTGCACTGCTGGGAGGGGAAGAGAT
AA031480.RC         108 CTGGGGGAGNC--CTN-TGCCGCCGCTGCACTGCTGGGAGGGGAAGAGAT
R52311               76 CTGGGGGAGCCTGCG-CTGCCGCCTTGGCANTGCTGGGAGGGGAAGAGAT
T33818               65 CTGGGGGAGCCTGCGGCTGCCGCCGTGNCACTGCTGGGAGGGGAAGAGAT
3087465              50 CTGAGGGAGCCTGCGGCTGCCGCCGCTGCACTGCTGGGAGGGGAAGAGAT
2525451              40 CTGGGGGAGCCTGCNGCTGCCGCCGCTGCACTGCTGGGAGGGGAAGANAT
C05296               34 CTGGGGGAGCCTGCGGCTGCCGCCGTGNCACTGCTGGGAGGGGAAGAGAT
C05172               31 CTGGGGGAGCCTGCGGCTGCCGCCGCTGCACTGCTGGGAGGGGAAGAGAT
2189372              30 CTGGGGGAGCCTGCGGCTGCCGCCGCTGCACTGCTGGGAGGGGAAGAGAT
HSC0CH081            26 CTGAGGGAGCCTGCGGCTGCCGCCGTNGCACTGCTGGGAGGGGAAGAGAT
1369095               3 CTGGGGGAGCCTGCGGCTGCCGNCGCTGNACTGCTGGGAGGGGAAGAGAT
2778468               1                                GCCGCTGCACTGCTGGGAGGGGAAGAGAT
T18925                1                                 CGCTGCACTGCTGGGAGGGGAAGAGAT
                        ++..+++++.++.+..++..++.+.....+.+..+.+++.+++++++.++
<consen02>         1039 CTGGGGGAGCCTGCGGCTGCCGCCGCTGCACTGCTGGGAGGGGAAGAGAT
```

FIGURE 4N

```
2760910          256 TTAGATCTGGACCAGGCTGTGGGTAGATGTGCAATAGAAA
2496946          253 TTAGATCTGGACCAG-CTGTGGGTAGATGTGCAATAGAAATA-CTAATTT
2223827          233 TTAGATCTGGACCAGGCTG
2226521          233 TTAGATCTGGACCAGGCTG
927569           227 TTAGATCTGGACCAGGCTGTGGGTAGATGTGCA
354028           178 TTAGATCTGGACCAGGCTGTGGGTAGATGTGCAATAGAAATAGCTAATTT
2277563          178 TTAGATCTGGACCAGGCTGTGGGTAAATGTGCAATAGAAATAGCTAATTT
1289239          177 TTAGATCTGGACCAGGCTGTGGGTAGATGTNCAATAGAAATAGCTAATTT
AA031480.RC      155 TTAGATCTGGACCAGGCTGTGGGTAGATGTGCAATAGAAATAGCTAATTT
R52311           125 TTAGATCTGGACCAGGCTGTGGGTAGATGTGCAATAGAAATAGCTAATTT
T33818           115 TTAGATCTGGACCAGGCTGTGGGTAGATGTGCAATAGAAATAGCTAATTT
3087465          100 TTAGATCTGGACCAGGCTGTGGGTAGATGTGCAATAGAAATAGCTAATTT
2525451           90 TTAGATCTGGACCAGGCTGTGGGTAGATGTGCAATAGAAATAGCTAATTT
C05296            84 TTAGATCTGGACCAGGCTGTGGGTAGATGTGCAATAGAAATAGCTAATTT
C05172            81 TTAGATCTGGACCAGGCTGTNGGTAGATGTGCAATAGAAATAGCTAATTT
2189372           80 TTAGATCTGGACCAGGCTGTGGGTAGATGTGCAATAGAAATAGCTAATTT
HSCOCH081         76 TTAGATCTGGACCAGGCTGTGGGTAGATGTGCAATAGAAATAGCTAATTT
1369095           53 TTAGATCTGGACCAGGCTGTGGGTAGATGTGCAATAGAAATAGCTAATTT
2778468           30 TTAGATCTGGACCAGGCTGTGGGTAGATGTGCAATAGAAATAGCTAATTT
T18925            28 TTAGATCTGGACCAGGCTGTGGGTAGATGTGCAATAGAAATAGCTAATTT
2885667            1     ATCTGGACCAGACTGTGGGTAGATGTGCAATAGAAATAGCTAATTT
H07079             1      TCTGGACC-GGNTGTGGGTAGATGTGCAATAGAAATAGCTAATTT
R15488             1                GCTGTGGGTAGATGTGCAATAGAAATAGCAAATTT
T80210             1                             GTCAANTAGAAATAGCTAATTT
H18569             1                              TCAATTAGAAATAGCTAATTT
                     ++++++++++++++..+++.++++.++++..+.++++++++++.+++++
<consen02>       1089 TTAGATCTGGACCAGGCTGTGGGTAGATGTGCAATAGAAATAGCTAATTT

2496946          301 ATTTCCCCAGGTGTGTGCTTTA
354028           228 ATTTCCCCAGGTGTGTGCTTTAGGCGTGGGCTGACCAGG-TTCTTCCTAC
2277563          228 ATTTCCCCAGGTGTGTGCTTTAGGCGTGGGCTGACCAGGC
1289239          227 ATTTCCCCAGGTGTGTGCTTTAGGCGTGGGCTGACCAGGCT
AA031480.RC      205 ATTTCCCCAAAAAA
R52311           175 ATTTCCCCAGGTNTGTGCTTTAGGCGTGGGCTGACCAGGCTTCTTCCNAC
```

FIGURE 4Q

```
T33818          165 ATTTCCCCAGGTGTGTGCTTTAGGCGTGGGCTGACCAGGCTTCTTNCTAC
3087465         150 ATTTCCCCAGGTGTGTGCTTTAGGCGTGGGCTGACCAGGCTTCTTCCTAC
2525451         140 AAAAAA
C05296          134 ATTTCCCCAGGTGTGTGCTTTAGGCGTGGGCTGACCAGGNTTCTTCCTAC
C05172          131 ATTTCCCCAGGTGTGTGCTTTAGGCGTGGGCTGACCAGTCTTCTTCCTAC
2189372         130 ATTTCCCCAGGTGTGTGCTTTAGGCGTGGGCTGACCAGGCTTCTTCCTAC
HSC0CH081       126 ATTTCCCCAGGTGTGTGCTTTAGGCGTGGGCTGANCAGGCTTCTTNCTAC
1369095         103 ATTTCCCCAGGTGTGTGCTTTAGGCGTGGGCTGACCAGGNTTCTTCCTAC
2778468          80 ATTTCCCCAGGTGTGTGCTTTAGGCGTGGGCTGACCAGGCTTCTTCCTAC
T18925           78 ATTTCCCCAGGTGTGTGCTTTAGGCGTGGGCTGCCCAGGCTTCTTCCTAC
2885667          47 ATTTCCCCAGGTGTGTGCTTTAGGCGTGGGCTGACCAGGCTTCTTCCTAC
H07079           45 ATTTCCCCAGGTGTGTGCTTTAGGCGTGGGCTGACCAGGCTTCTTCCTAC
R15488           36 ATTTCCCCAGGTGTGTGCTTTAGGCGTGGGCTGACCAGGCTTCTTCCTAC
T80210           23 ATTTCCCCAGGTGTGTGCTTTAGGCGTGGNCTGACCAGGCTTCTTCCTAC
H18569           22 ATTTCCCCAGGTGTGTGCTTTAGGCGTGGGCTGACCAGGCTTCTTCCTAC
045844            1        CCAGGTGTGTGCTTTAGGCGTGGGCTGACCAGGCTTCTTCCTAC
1704942           1            GTGTGTGCTTTAGGCGTGGGCTGACCAGGCTTCTTCCTAC
877050            1                    CTTTAGGCGTGGGCTGACCAGGCTTCTTCCTAC
865207            1                         GGCGTGGGCTGACCAGGCTTCTTCCTAC
833987            1                         GGCGTGGGCTGACCAGGCTTCTTCCTAC
2500233           1                              GGGCTGACCAGGCTTCTTCCTAC
N86855            1                                GCTGACCAGGCTTCTTCCTAC
1405577           1                                    ACCAGGCTTCTTCCTAC
2056655           1                                      CCAGGCTTCTTCCTAC
945395            1                                      CCAGGCTTGTTNCTAC
354502            1                                        AGGCTTCTTCCTAC
                    +.....+++.....++++++++++++++.+++..+++..++.++.+.++
<consen02>      1139 ATTTCCCCAGGTGTGTGCTTTAGGCGTGGGCTGACCAGGCTTCTTCCTAC

354028          277 ANCTTCTTCCCAGTAAGTTTCCCCTCTGGCTT-NACANCATGAGGTGTTG
R52311          225 ATCTTCTTCCCAGTAAGTTTCCCCTCTGGCTT-GACAGCATGAGGTGTTN
T33818          215 ATCTTCTTCCCAGTAAGTTTCCCCTCTGGCTT-GACAGCATGAGGTGTTG
3087465         200 ATCTTCTTCCCAGTAAGTTTCCCCTCTGGCTT-GACAGCATGAGGTGTTG
C05296          184 ATCTTCTTCCCAGTAAGTTTCCCCTCT
```

FIGURE 4P

```
C05172          181 ATCTTCTTCCCANTAAGTTTCCCCTCTGGCTT-GACAGCATGAGGTGTTG
2189372         180 ATCTTCTTCCCAGTAAGTTTCCCCTCTGGCTT-GACAGCATGAGGTGTTG
HSC0CH081       176 ATCTTCTTGCCAGTANGNTTCCCCTCTGGCTT-GACAGCATGAGGTGTTG
1369095         153 ATCTTNTTCCCAGTAAGTTTCCCCTCTGGCTT-GACAGNATGAGGTGTTG
2778468         130 ATCTTCTTCCCAGTAAGTTTCCCCTCTGGCTT-GACAGCATGAGGTGTTG
T18925          128 ATNTCCGTCCCNGTAAGTTTCCCCTCTAGCGA-AAACAGAATAAGGTG
2885667          97 ATCTTCTTCCCAGTAAGTTTCCCCTCTGGCTT-GACAGCATGAGGTGTTG
H07079           95 ATCTTCTTCCCAGTAAGTTTCCCCTCTGGCTT-GACAGCATGAGGTGTTG
R15488           86 ATCTTCTTCCCAGTAAGTTTCCCCTCTGGCTT-GACAGCATGAGGTGTTG
T80210           73 ATCTTCTTCCCAGTAAGTTTCCCCTCTGGCTTTGACAGCATGAGGTGTTG
H18569           72 ATCTTCTTCCCAGTAAGTTTCCCCTCTGGCTT-GACAGCATGAGGTGTTG
045844           45 ATCTTCTTCCCAGTAAGTTTCCCCTCTGGCTT-GACAGCATGAGGTGTTG
1704942          41 ATCTTCTTCCCAGTAAGTTTCCCCTCTGGCTT-GACAGCATGAGGTGTTG
877050           34 ATCTTCTTCCCAGTAAGTTTCCCCTCTGGCTT-GACAGCATGAGGTGTTG
865207           29 ATCTTCTTCCCAGTAAGTTTCCCCTCTGGCTT-GACAGCATGAGGTGTTG
833987           29 ATCTTCTTCCCAGTAAGTTTCCCCTCTGG-TT-GACAGCATGAGGTGTTG
2500233          24 ATCTTCTTCCCAGTAAGTTTCCCCTCTGGCTT-GACAGCATGAGGTGTTG
N86855           22 ATCTTCTTCCCAGTAAGTTTCCCCTCTGGCTT-GACAGCATGAGGTGTTG
1405577          18 ATCTTCTTCCCAGTAAGTTTCCCCTCTGGCTT-GACAGCATGAGGTGTTG
2056655          17 ATCTTCTTCCCAGTAAGTTTCCCCTCTGGCTT-GACAGCATGAGGTGTTG
945395           17 ATCTTNTTNCCAGTAAGTTTNCCNTCTGGCTT-GACAGCATGAGGTGTTG
354502           15 ATCTTCTTCCCAGTAAGTTTCCCCTCTGGCTT-GACAGCATGAGGTGTTG
681862            1   CTTCTTCCCAGTAAGTTTCCCCTCTGGCTT-GACAGCATGAGGTGTTG
2547081           1                    TTCCCCTCTGGCTT-GACAGCATGAGGTGTTG
381195            1                       CCCTCTGGCTT-GACAGCATGAGGTGTTG
947376            1                                                  G
                    +..+...+.++..++.+.++.++.+++.++.. .+....+..+.+...+.
<consen02>      1189 ATCTTCTTCCCAGTAAGTTTCCCCTCTGGCTT GACAGCATGAGGTGTTG
```

FIGURE 4Q

354028	326	TGCAATTGTTCAACTTNCCCCCANGNTTGTTTTCCANGGTTCANANTCTT
R52311	274	TGCATTTGTTCAGCT--CCCCCAGGCT-GTTCTCCAGGCTTCACAGTCT-
T33818	264	TGCATTTGTTCAGCT--CCCCCAGGCT-GTTCTCCAGG
3087465	249	TGCATT-GTTCAGCT--CCCCCAGGCT-GTTCTCCAGGCTTC
C05172	230	TGCATTTTTTCAGCT--CCCCCAGGCT-GTTCTCCAGGCTTCACAGTCT-
2189372	229	TGCATTTGTTCAGCT--CCCCCAGGCT-GTTCTCCAGGCTTCACAGTCT-
HSC0CH081	225	TGCATTTGTTCAGCT--CCCCCAGGCT-GTTCTCCAGGCTTCACAGTCT-
1369095	202	TGCATTTGTTCAG
2778468	179	TGCATTTGTTCAGCT--CCCCCAGGCT-GTTCTCCAGGCTTCACAGTCT-
2885667	146	TGCATTTGTTCAGCT--CCCCCAGGCT-GTTCTCCAGGCTTCACAGTCT-
H07079	144	TGCATTTGTTCAGCT--CCCCCAGGCT-GTTCTCCAGGCTTCACAGTCT-
R15488	135	TGCATTTGTTCAGCT--CCCCCAGGCT-GTTCTCCAGGCTTCACAGTCT-
T80210	123	TGCATTTGTTCAGCT--CCCCCAGGCT-GTTCTCCAGGCTTCACAGTCT-
H18569	121	TGCATTTGTTCAGCT--CCCCCAGGCT-GTTCTCCAGGCTTCACAGTCT-
045844	94	TGCATTTTTNCAGNT--NCCCCAGGCT-GTTCTCCAGGCTTNANAGTCT-
1704942	90	TGCATTTGTTCAGCT--CCCCCAGGCT-GTTCTCCAGGCTTCACAGTCT-
877050	83	TGCATTTGTTCAGCT--CCCCCAGGNT-GTTCTCCAGGCTTCACAGTCT-
865207	78	TGCATTTGTTCAGCT--CCCCCAGGCT-GTTCTCCAGGCTTCACAGTCT-
833987	77	TGCATTTGTTNAGCT--CCCCCAGGCT-GTTCTNCAGGCTTCACAGTCT-
2500233	73	TGCATTTGTTCAGCT--CCCCCAGGCT-GTTCTCCAGGCTTCACAGTCT-
N86855	71	TGCATTTGTTCAGCT--CCCCCAGGCT-GTTCTCCAGGCTTCACAGTCT-
1405577	67	TGCATTTGTTCAGCT--CCCCCAGGCT-GTTCTCCAGGCTTCACAGTCT-
2056655	66	TGCATTTGTTCAGCT--CCCCCAGGCT-GTTCTCCAGGCTTCACAGTCT-
945395	66	TGCATTTGTTCAGNT--NCCCCAGGGT-GTTNTNCAGGGTTCACAGTNT-
354502	64	TGCATTTGTTCAGCT--CCCCCAGGCT-GTTCTCCAGGCTTCACAGTCT-
681862	48	TGCATTTGTTCAGCT--CCCCCAGGCT-GTTNTCCAGGNTTCACAGTCT-
2547081	32	TGCATTTGTTCAGCT--CCCCCAGGCT-GTTCTCCAGGCTTCACAGTCT-
381195	29	TGCATTTGTTCAGCT--CCCCCAGGCT-GTTCTCCAGGCTTCACAGTCT-
947376	2	TGCATTTNTTCAGNT--CCCCNAGGCT-GTTCTCCAGGCTTNACAGTCT-
1351726	1	CGGCTCGAGCTCGAG--CCGATTCGGC-TCGAGCCAGGCTTCACAGTCT-
624084	1	ATTTGTTCAGCT--CCCCCAGGCT-GTTCTCCAGGNTTCACAGTCT-
622154	1	ATTTGTTCAGCT--CCCCCAGGNT-GTTCTNCAGGNTTCACAGTCT-
1727802	1	CTCCAGGCTTCACAGTCT-
HSC3MH091	1	CTCCAGGCTTCACAGTCT-

FIGURE 4R

```
                .+.............  .+.....+.. ......++.+.++.+.+.+.+
<consen02>     1238 TGCATTTGTTCAGCT  CCCCCAGGCT GTTCTCCAGGCTTCACAGTCT

354028          376 NGTGCTTTGG-AGAGTC-ANGCAGGGTTAAACTNCAGGAGCANTTTTCCA
R52311          320 TGTGCTTGGG-AGAGTC-AGGCAGGGTTAAACTGCAGGAGCAGTTTGCCA
C05172          276 GGTGCTTGGG-AGAGTC-AGGCAGGGTTAAACTNCAGGAGCAGTTT
2189372         275 GG
HSC0CH081       271 GGTGCTTGG
2778468         225 GGTGCTTGGG-AGAGTC-AGGCAGGGTTA
2885667         192 GGTGCTTGGG-AGAGTC-AGGCAGGGTTAAACTGCAGGAGCAGTTTGCCA
H07079          190 GGTGCTTGGG-AGAGTC-AGGCAGGGTTAAACTGCAGGAGCAGTTTGCCA
R15488          181 GGTGCTTGGG-AGAGTC-AGGCAGGGTTAAACTGCAGGAGCA-TTTGCCA
T80210          169 GGTGCTTGGG-AGAGTC-AGGCAGGGTTAAACTGCAGGAGCAGTTTGCCA
H18569          167 GGTGCTTGGG-AGAGTC-AGGCAGGGTTAAACTGCAGGAGCAGTTTGCCA
045844          140 GGTGCTTGGG-AGAGTC-AGGCAGNGTTAAACTGCAGGAGCAGTTTGCCA
1704942         136 GGTGCTTGGG-AGAGTC-AGGCAGGGTTAAACTGCAGGAGCAGTTTGCCA
877050          129 GGTGCTTGGG-AGAGTC-AGGCAGGGTTAAACTGCAGGAGCAGTTTGCCA
865207          124 GGTGCTTGGG-AGAGTC-AGGCAGGGTTAAACTGCAGGAGCAGTTTGCCA
833987          123 GGTGCTTGGG-AGAGTC-AGGCAGGGTTAAACTGCAGGAGCAGTTTGCCA
2500233         119 GGTGCTTGGG-AGAGTC-AGGCAGGGTTAAACTGCAGGAGCAGTTTGCCA
N86855          117 GGTGCT-GGG-AGAGTC-AGGCAGGGTTAAACTGCAGGAGCAGTTTGCCA
1405577         113 GGTGCTTGGG-AGAGTC-AGGCAGGGTTAAACTGCAGGAGCAGTTTGCCA
2056655         112 GGTGCTTGGG-AGAGTC-AGGCAGGGTTAAACTGCAGGAGCAGTTTGCCA
945395          112 GGTGCTTGGG-AGAGTC-ANGCAGGGTTAAACTGCAGGAGCAGTTTGCCA
354502          110 GGTGCTTGGG-AGAGTC-AGGCAGGGTTAAACTGCAGGAGCAGTTTGCCA
681862           94 GGTGCTTGGG-AGAGTC-AGGCAGGGTTAAACTGCAGGAGCAGTTTGCCA
2547081          78 GGTGCTTGGG-AGAGTC-AGGCAGGGTTAAACTGCAGGAGCAGTTTGCCA
381195           75 GGTGCTTGGG-AGAGTC-AGGCAGGGTTAAACTGCAGGAGCAGTTTGCCA
947376           48 NGTGCTTNGG-AGAGTC-AGGCAGGNTTAAACTGCAGGAGCAGTTTGCCA
1351726          47 GGTGCTTGGG-AGAGTC-AGGCAGGGTTAAACTGCAGGAGCAGTTTGCCA
624084           44 GGTGCTTGGG-AGAGTC-AGGNAGGGTTAAACTGCAGGAGCAGTTTGCCA
622154           44 GGTGCTTGGG-AGAGTC-AGGCAGGGTTAAACTGCAGGAGCAGTTTGCCA
1727802          19 GGTGCTTGGG-AGAGTC-AGGCAGGGTTAAACTGCAGGAGCAGTTTGCCA
HSC3MH091        19 GGTGCTTGGG-AGAGTC-ACGCAGGGTTAAACTGCAGGAGCAGTTTGCCA
```

FIGURE 4S

```
2156323          1    GCTTGGGGAGAGTCGAGGCAGGGTTAAA-TGCAGGAGCAGTTTGCCA
463503           1         GG-AGAGTC-AGGCAGGGTTAAACTGCAGGAGCAGTTTGCCA
T78829           1                 C-ACGCAGGGTTAAACTGCAGGAGCAGTTTGCCA
HSC37E101        1                     GCAGGGTTAAACTGCAGGAGCA-TTTGCCA
465657           1                          GTTAAACTNCAGGAGCAGTTTGCCA
N36746           1                                    AGTTTGCCA
T65270           1                                         CCA
                    .++++++.++ ++++++ +.+.++..+++++++.++++++++.+++.+++
<consen02>    1284 GGTGCTTGGG AGAGTC AGGCAGGGTTAAACTGCAGGAGCAGTTTGCCA

354028         424 NCC-T-GTTCCAGATTATT-GGCTTCTTT-CCTCT
R52311         368 CCCCT-GT-CCAGATTATTTGGCTGCTTTGCC
2885667        240 CCCCT-GT-CCAGATTATT-GGCTGCTTTGCCTCTACCAGTTGG-CAGAC
H07079         238 CCCCT-GT-CCAGATTATT-GGCTGCTTTGCCTCTACCAGTTGGGCAGAC
R15488         228 CCCCT-GT-CCAGATTATT-GGCTGCTTTGC-TCTACCAGTTGG-CAGAC
T80210         217 CCCCT-GT-CCAGATTATT-GGCTGCTTTGCCTCTACCAGTTGG-CAGAC
H18569         215 CCCCT-GT-CCAGATTATT-GGCTGCTTTGCCTCTACCAGTTGG-CAGAC
045844         188 CCCCT-GT-CCAGATTATT-GGCTGCTTTGCCTCTACCANTTGG-CAGAC
1704942        184 CCCCT-GT-CCAGATTATT-GGCTGCTTTGCCTCTACCAGTTGG-CAGAC
877050         177 CCCCT-GT-CCAGATTATT-GGCTGCTTTGCCTCTACCAGTTGG-CAGAC
865207         172 CCCCT-GT-CCAGATTATT-GGCTGCTTTGCCTCTACCAGTTGG-CAGAC
833987         171 CCCCT-GT-CCAGATTATT-GGCTGCTTTGC-TCTACCAGTTGG-CAGAC
2500233        167 CCCCT-GT-CCAGATTATT-GGCTGCTTTGCCTCTACCAGTTGG-CAGAC
N86855         164 CCCCT-GT-CCAGATTATT-GGCTGCTTTGCCTCTACCAGTTGG-CAGAC
1405577        161 CCCCT-GT-CCAGATTATT-GGCTGCTTTGCCTCTACCAGTTGG-CAGAC
2056655        160 CCCCT-GT-CCAGATTATT-GGCTGCTTTGCCTCTACCAGTTGG-CAGAC
945395         160 CCCNT-GT-CCAGATTATT-GGNTGNTTTGNCTGTACCAGTTGG-CAGAC
354502         158 CCCCT-GT-CCAGATTATT-GGCTGCTTTGCCTCTACCAGTTGG-CAGAC
681862         142 CCCCT-GT-CCAGATTATT-GGCTGCTTTGCCTCTACCAGTTGG-CAGAC
2547081        126 CCCCT-GT-CCAGATTATT-GGCTGCTTTGCCTCTACCAGTTGG-CAGAC
381195         123 CCCCT-GT-CCAGATTATT-GGCTGCTTTGCCTCTACCAGTTGG-CAGAC
947376          96 CCCCT-GT-CCAGATTATT-GGCTGCTTNGCCTCTACCAGTTGG-CAGAC
1351726         95 CCCCT-GT-NCAGATTATT-NNCTGCTTTGCCTCTACCAGTTGG-CAGAC
624084          92 CCCCT-GT-CCAGATTATT-GGCTGCTTTGCCTCTACCAGTTGG-CAGAC
```

FIGURE 4T

```
622154          92 CCCCT-GT-CCAGATTATT-GGCTGCTTTGCCTCTACCAGTTGG-CAGAC
1727802         67 CCCCT-GT-CCAGATTATT-GGCTGCTTTGCCTCTACCAGTTGG-CAGAC
HSC3MH091       67 CCCCT-GT-CCAGATTATT-GGCTGCTTTGCCTCTACCAGTTGG-CAGNC
2156323         47 CCCCT-GT-CCAGATTATT-GGCTGCTTTGNCTCTACCAGTTGG-CAGAC
463503          41 CCCCT-NT-CCAGATTATT-GGCTGCTTTGCCTCTACCAGNTGG-CAGNC
T78829          34 CCCCTTGT-CCAGATTATT-GGCTGCTTTNC-TCTACCAGTTGG-CAGAC
HSC37E101       30 CCCCT-GT-CCAGATTATT-GGCTGCTTTGCCTCTACCAGTTGG-CAGNC
465657          26 CCCCT-GT-CCAGATTATT-GGCTGCTTTGCCTCTACCAGTTGG-CAGAC
N36746          10 CCCCT-GT-CCAGATTATT-GGCTGCTTTGCCTCTACCAGTTGG-CAGAC
T65270           4 CCCCT-GT-CCAGATTATT-GGCTGCTTTNC-TCTACCAGTTGG-CAGAC
1503017          1  CCCT-GT-CCAGATTATT-GGCTGCTTTGCCTCTACCAGTTGG-CAGAC
1370618          1          AGATTATT-GGCTGCTTTGCCTCTACCAGTTGG-CAGAC
712299           1                   GCCTNTACCAGTTGG-CAGAC
2899383          1                              TTGG-CAGAC
2666235          1                           GTTGG-CAGAC
2846187          1                                 GAC
                    .++.+ .+ .++++++++++ ...+..++...++.++++..+++ +++.+
<consen02>      1332 CCCCT GT CCAGATTATT GGCTGCTTTGCCTCTACCAGTTGG CAGAC

2885667         286 A
H07079          285 AGCCGTTTGTTCTACATGGGCTTTGATAA-TTGT-TTGAG-GGGGAGGGA
R15488          273 AGC-GTTTGTTCTACATGG-CTTTGATAA-TTGT-TTGAG-GG--AGG-A
T80210          263 AGCCGTTTGTTCTACATGG-CTTTGATAA-TTGT-TTGAG-GGG-AGG-A
H18569          261 AGCCGTTTGTTCTACATGG-CTTTGATAA-TTGT-TTGAG-GGG-AGG-A
045844          234 AGNCGTTTGTTCTACATGG-CTTTT
1704942         230 A
877050          223 AGCCGTTTGTTCTACATGG-CTT
865207          218 AGCCGTTTGTTCTACATGG-CTTTGATAA-TTGT-TTGAG-GGG-AGG-A
833987          216 AGCCGTTTGTTCTACATGG-CTTTGATAA-TTGT-TTGAG-GGG-AGG-A
2500233         213 AGCCGTTTGTTCTACATGG-CTTTGATAA-TTG
N86855          210 AGCCGTTTGTTCTACATGG-CTTTGATAA-TTGT-TTGGA-GGA-AGG-A
1405577         207 AGCCGTTTGTTCTACATGG-CTTTGATAA-TTGT-TTGAG-GG
2056655         206 AGCCGTTTGTTCTACATGG-CTTTGATAA-TTGT-TTGAG-GGG-AGG-A
945395          206 AGCCGTTTGTTNTACATGG-NTTTGATAA-TTGT-TTGAG-GGG-AGG-A
```

FIGURE 4U

```
354502          204 AGCCGTTTGTTCTACATGG-CTTTGATAA-TT
681862          188 AGCCGTTTGTTCTACATGG-NTTTGATAA-TTGT-TTG
2547081         172 AGCCGTTTGTTCTACATGG-CTTTGATAA-TTGT-TTGAG-GGG-AGG-A
381195          169 AGCCGTTTGTTCTACATGG-CTTTGATAA-TTGT-TTGAG-GGG-AGG-A
947376          142 AGCCGTTTGTTCTACATGG-CTTTGATAA-TTGT-TTGAG-GGG-AGG-A
1351726         141 AGCCGTTNGTTCTACATGG-CTTTGATAA-TTGT-TTGAG-GGG-AGG-A
624084          138 AGCCGTTTGTTCTACATGG-NTTTGATAA-TTGT-TTGAG-GGG-AGG-A
622154          138 AGCCGTTTGTTCTACATGG-NTTTGATAA-TTGT-TTGAG-GGG-AGG-A
1727802         113 AGCCGTTTGTTCTACATGG-CTTTGATAA-TTGT-TTGAN-GGG-AGG-A
HSC3MH091       113 AGCCGTTTGTTCTACATGG-CTTTGATAA-TTGT-TTGNG-GGG-AGG-A
2156323          93 AGCCGTT-GTCCNACATGG-CTTTGANAANTTNCCNTCAG-GGG-ANG-A
463503           87 AGCCGTTTGTTCTACATGG-CTTTGATAA-TTGN-TTGNG-GGG-AGG-A
T78829           80 AGC-GTTTGTTCTACATGG-CTTTGATAA-TTGT-TTGAG-GGG-AGG-A
HSC37E101        76 AGCCGTTTGTTCTACATGG-CTTTGATAA-TTGT-TTGAG-GGG-AGG-A
465657           72 AGCCGTTTGTTCTACATGG-CTTTGATAA-TTGT-TTGAG-GGG-AGG-A
N36746           56 AGCCGTTTGTTCTACATGG-CTTTGATAA-TTGT-TTGAG-GGG-AGG-A
T65270           49 AGC-GTTTGTTCTACATGG-CTTTGATAA-TTGT-TTGAG-GGG-AGG-A
1503017          46 AGCCGTTTGTTCTACATGG-CTTTGATAA-TTGT-TTGAG-GGG-AGG-A
1370618          38 AGCCGTTTGTTCTACATGG-CTTTGATAA-TTGT-TTGAG-GGG-AGG-A
712299           21 AGCCGTTTGTTCTACATGG-NTTTGATAA-TTGT-TTGGG-GGG-AGG-A
2899383          10 AGCCGTTTGTTCTACATGG-CTTTGATAA-TTGT-TTGNCCGGG-AGG-A
2666235          11 AGCCGTTTGTTCTACATGG-CTTTGATAA-TTGT-TTGAN-GGG-AGG-A
2846187           4 AGCCGTTTGTTCTACATGG-CTTTGATAA-TTGT-TTGAN-GGG-AGG-A
1358133           1                                                GG-AGG-A
1349616           1                                                   GG-A
                    ++.++++.++...++++++ .+++.+.++ ++.. .+... ++. +.+ +

<consen02>      1378 AGCCGTTTGTTCTACATGG CTTTGATAA TTGT TTGAG GGG AGG A
```

FIGURE 4V

```
H07079       332 GATGGGAAACAATGTGGGAGTCTCCCTCT-GATTGGTTTTGGGGGAAATG
R15488       315 GATGG-AA-CAATGTGGAGT
T80210       307 GATGG-AAACAATGTGGGAGTCTCCCTCT-GATTGGTTTTGGGG-AAATG
H18569       305 GATGG-AAACAATGTGGGAGTCTCCCTCT-GATTGGTTTTGGGG-AAATG
865207       262 G
833987       260 GATGG-AAAC
N86855       254 GATGG-AAACAATGTGG-AGTCTCCCTCTTGAT-GGTTT-GGGG-AAAGT
2056655      250 GATGG-AAACAATGTGG-AGTCTCCCT
945395       250 GATGG-AAACAATGTGG-AGTNTGCCTGT-GATTGGT
2547081      216 GATGG-AAACAATGTGG-AGTCTCCCTCT-G
381195       213 GATGG-AAACAATGTGG-AGTCTCCCTCT-GATTGGTTTTGGGG-AAATG
947376       186 GATGG-AAACAATGTGG-AGTCTCCCTCT-GATTGGTTTTGGGG-AAATN
1351726      185 GATGG-AAACAATGTGG-AGTCTCCCTCT-GATTGGTTTTGGGG-AAATG
624084       182 GATGG-AAACAATGTGG-AGTCTCCCTCT-GATTGGTTTTGGGG-AAATG
622154       182 GATGG-AAACAATGTGG-AGTCTCCCTCT-GATTGGTTTTGGGG-AAATG
1727802      157 GATGG-AAACAATGTGG-AGTCTCCCTCT-GATTGGTTTTGGGG-AAATG
HSC3MH091    157 GATGG-AAACAATGTGG-AGTCTCCCTCT-GATTGGTTTTGGGG-AAATG
2156323      138 GATGG-AAANNANGNGGAGCNCNGGCTTT-TCCCCGTTGGG
463503       131 GATGG-AAACAATGTGG-AGTCTCCCTCT-GATTGGTTTTGGGG-AAATG
T78829       123 GATGG-AAACAATGTGG-AGTCTCCCTCT-GATTGGTTTTGGGG-AAATG
HSC37E101    120 GATGG-AAACAATGTGG-AGTCTCCCTCT-GATTGGTTTTGGGG-AAATG
465657       116 GATGG-AAACAATGTGG-AGTCTCCCTCT-GATTGGTTTTGGGG-AAATG
N36746       100 GATGG-AAACAATGTGG-AGTCTCCCTCT-GATTGGTTTTGGGG-AAATG
T65270        92 GATGG-AAACAATGTGG-AGTCTCCCTCT-GATTGGTTTTGGGG-AAATG
1503017       90 GATGG-AAACAATGTGG-AGTCTCCCTCT-GATTGGTTTTGGGG-AAATG
1370618       82 GATGG-AAACAATGTGG-AGTCTCCCTCT-GATTGGTTTTGGGG-AAATG
712299        65 GATGG-AAACAATGTGG-AGTCTTCCTCT-GATTGGTTTTTGGG-AAATG
2899383       55 GATGG-AAACAATGTGG-AGTCTCCCTCT-GATTGGTTTTGGGG-AAATG
2666235       55 GATGG-AAACAATGTGG-AGTCTCCCTCT-GATTGGTTTTGGGG-AAATG
2846187       48 GATGG-AAACAATGTGG-AGTCTCCCTCT-GATTGGTTTTGGGG-AAATG
1358133        7 GATGG-AAACAATGTGG-AGTCTCCCTCT-GATTGGTTTTGGGG-AAATG
1349616        4 GATGG-AAACAATGTGG-AGTCTCCCTCT-GATTGGTTTTGGGG-AAATG
2305359        1 GATGG-AAACAATGTGG-AGTCTCCCTCT-GATTGGTTTTGGGG-AAATG
2149106        1    GG-AAACAATGTGG-AGTCTCCCTCT-GATTGGTTTTGGGG-AAATG
```

FIGURE 4W

```
1984534          1    GG-AAACAATGTGG-AGTCTCCCTCT-GATTGGTTTTGGGG-AAATG
353611           1                GTGGAGTCTCCCTCT-GATTGGTTTTGGGG-AAATG
052714           1                          CNCT-GATTGGTTTTGGGG-AAATN
AA041443         1                          CTCT-GATTGGTTTTGGGG-AAATG
AA090223         1                                              CG-AAATG
2152542          1                                              GG-AAATG
                   +++++ +++..+.+..+........+..+ .....+++...+.+ +++..
<consen02>    1422 GATGG AAACAATGTGGGAGTCTCCCTCT GATTGGTTTTGGGG AAATG

H07079         381 TGGGAGAAGGAGTNCCCTGCTTTTGCAAACATCAA-CCTGGGCAAAANTG
T80210         354 TGGGAGAAG-AGTGCCCTGCTTTTGCAAACATCAA-CCTGGGCAAAAATG
H18569         352 TGGGAGAAG-AGTGCCCTGCTTTTGCAAACATCAA-CCNGGGCAAAAATG
N86855         299 GG
381195         259 TGG-AGAAG-AGTG
947376         232 TGG-AGAAG-AGTGCCCTT-TTT-ANAAAAATCAA-NCCGGCAAAAAATG
1351726        231 TGGA
624084         228 TGG-AGAAG-AGTGCC
622154         228 TGG-AGAAG-AGTGCCCTGCTTT-NCAAA
1727802        203 TGGAGAAC
HSC3MH091      203 TGG-AGAAG-AGTGCCCTGCTTT-GCAAACATCAA-CCTGG-CAAAAATG
463503         177 TGG-AGAAG-AGTGCCCTGCTTT-GCAAACATCAA-CCTGG-CAAAANTG
T78829         169 TGG-AGAAG-AGTGCCCTGCTTT-GCAAACATCAA-CCTGG-CAAAAATG
HSC37E101      166 TGG-AGAAG-AGTGCCCTGCTTT-GCAAACATCAA-CCTGG-CAAAAATG
465657         162 TGG-AGAAG-AGTGCCCTGCTTT-GCAAACATCAA-CCTGGN-AAAAATG
N36746         146 TGG-AGAAG-AGTGCCCTGCTTT-GCAAACATCAA-CCTGG-CAAAAATG
T65270         138 TGG-AGAAG-AGTGCCCTACTTT-NCAAACATCAA-CCTNGC
1503017        136 TGG-AGAAG-AGTGCCCTGCTTT-GCAAACATCAA-CCTGG-CAAAAATG
1370618        128 TGG-AGAAG-AGTGCCCTGCTTT-GCAAACATCAA-CCTGG-CAAAAATG
712299         111 TGG-AGAAG-AGTNCCCTGCTTT-GCAAACATCAA-CCTGG-CAAAAATG
2899383        101 TGG-AGAAG-AGTGCCCTGCTTT-GCAAACATCAA-CCTGG-CAAAAATG
2666235        101 TGG-AGAAG-AGTGCCCTGCTTT-GCAAACATCAA-CCTGG-CAAAAATG
2846187         94 TGG-AGAAG-AGTGCCCTGCTTT-GCAAACATCAA-CCTGG-CAAAAATG
1358133         53 TGG-AGAAG-AGTGCCCTGCTTT-GCAAACATCAA-CCTGG-CAAAAATG
1349616         50 TGG-AGAAG-AGTGCCCTGCTTT-GCAAACATCAA-CCTGG-CAAAAATG
```

FIGURE 4X

```
2305359          47 TGG-AGAAG-AGTGCCCTGCTTT-GCAAACATCAA-CCTGG-CAAAAATG
2149106          44 TGG-AGAAG-AGTGCCCTGCTTT-GCAAACATCAA-CCTGG-CAAAAATG
1984534          44 TGG-AGAAG-AGTGCCCTGCTTT-GCAAACATCAA-CCTGG-CANAAATG
353611           35 TGG-AGAAG-AGTGCCCTGCTTT-GCAAACATCAA-CCTGG-CAAAAATG
052714           24 TGG-AGAAG-AGTGCCCTGCTTT-GCAAACATCAA-CCTGG-CAAAAATG
AA041443         24 TGG-AGAAG-AGTGCCCTGCTTT-GCAAACATCAA-CCTGG-CAAAAATG
AA090223          8 TGG-AGAAG-AGTGCCCTGCTTT-GCAAACATCAA-CCTGG-CAAAAATG
2152542           8 TGG-AGAAG-AGTGCCCTGCTTT-GCAAACATCAA-CCTGG-CAAAAATG
N26110.RC         1     GAAAG-AGTCCCCTGCTTT-GCAAACATCAAACCTGGCAAAAAATN
857436            1   GGAGAAG-AGTGCCCTGCTTT-GCAAACATCAA-CCTGGGCAAAAATG
1377826           1   GGAGAAG-AGTGCCCTGCTTT-GCAAACATCAA-CCTGG-CAAAAATG
                    .++...+.+ +++.++++.++++ ..+++.+++++ .+..+..+.++.+.
<consen02>     1469 TGGGAGAAG AGTGCCCTGCTTT GCAAACATCAA CCTGGGCAAAAATG

H07079          430 CAACAA
T80210          402 CAACAAA-T-GAATTTTTCC-ACGNAGTT
H18569          400 CAACCAA-TTGAATTTTTCC-NCGCANTTTCNTTTCC-CTTGGGCA-TAA
947376          277 CAACAAA-T-GTNTTTT-CC-GNGGTGTT-N-TTNCC-NG-GG
HSC3MH091       248 CAACAAA-T-GAATTTT-CC-ACGCAGTT-C-TTTCC-AT-GGGCA-TA-
463503          222 CAACAAA-T-GNT
T78829          214 CAACAAA-T-GAATTTT-CC-ACGCAGTT-C-TTTCC-AT-GGGCA-TA-
HSC37E101       211 CAACAAA-T-GAATTTT-CC-ACGCA-TT-C-TTTCC-AT-GGGCA-TA-
465657          207 CAACAAA-T-GAAT
N36746          191 CAACAAA-T-GAATTTT-CC-ACGCAGTT-C-TTTCC-AT-GGGCA-TA-
1503017         181 CAACAAA-T-GAATTTT-CC-ACGCAGTT-C-TTTCC-AT-GGGCA-TA-
1370618         173 CAACAAA-T-GAATTTT-CC-ACGCAGTT-C-TTTCC-AT-GGGCA-TA-
712299          156 CAACAAA-T-GAATTTT-NC-ACGNAGTT-C-TTTCC-AT-GGGCA-TA-
2899383         146 CAACAAA-T-GAATTTT-CC-ACGCAGTT-C-TTTCC-AT-GGGCA-TA-
2666235         146 CAACAAA-T-GAATTTT-CC-ACGCAGTT-C-TTTCC-AT-GGGCA-TA-
2846187         139 CAACAAA-T-GAATTTT-CC-ACGCAGTT-C-TTTCC-AT-GGGCA-TA-
1358133          98 CAACAAAAT-GAATTTT-CC-ACGCANTT-C-TTTCC-AT-GGGGA-TA-
1349616          95 CAACAAA-T-GAATTTT-CC-ACGCAGTT-C-TTTCC-TT-GGGCA-TA-
2305359          92 CAACAAA-T-GAATTTT-CC-ACGCAGTT-C-TTTCC-AT-GGGCA-TA-
2149106          89 CAACAAA-T-GAATTTT-CC-ACGCAGTT-C-TTTCC-AT-GGGCA-TA-
```

FIGURE 4X

```
1984534          89 CAACAAA-T-GAATTTT-CC-ACGCAGTT-C-TTTCC-AT-GGGCA-TA-
353611           80 CAACAAA-T-GAATTTT-CC-ACGCAGTT-C-TTTCC-AT-GGGCA-TA-
052714           69 CAACAAA-T-GANTTTC-CC-ACGCAGTT-C-TTTCC-AT-GGGCA-TA-
AA041443         69 CAACAAA-T-GAATTTT-CC-ACGCAGTT-C-TTTCC-AT-GGGCA-TA-
AA090223         53 CAACAAA-T-GAATTTT-CC-ACGCAGTT-C-TTTCC-AT-GGGCA-TA-
2152542          53 CAACAAA-T-GAATTTT-CC-ACGCAGTT-C-TTTCC-AT-GGGCA-TA-
N26110.RC        45 CAACAAAAT-GAATTTT-CCNACGCAGTT-C-TTTCCCAT-GGGCAATA-
857436           46 CAACAAA-T-GAATTTT-CC-ACGNAGTT-C-TTTCC-AT-GGGCA-TA-
1377826          45 CAACAAA-T-GAATTTT-CC-ACGCAGTT-C-TTTCC-AT-GGGCA-TA-
                    ++++.++ + +..+++. .+ ..+...++ . ++.++ .. +++.+ ++
<consen02>     1516 CAACAAA T GAATTTT CC ACGCAGTT C TTTCC AT GGGCA TA

H18569          446 GGTAAGCTGTTGCCCTTCAAGCTGTTTTGCCAGAAGNAATTGTTCTGGTT
HSC3MH091       288 GGTAAGCTGT-GCC-TTCA-GCTGTTGCAGATGAAATGTTCTGTTCACCC
T78829          254 GGTAAGCTGT-GCC-TTCA-GCTGTTGCAGATGAAATGTTCTGTTCACCC
HSC37E101       250 GGTAAGCTGT-GCC-TTCA-GCTGTTGCAGATGAAATGTTCTGTTCACCC
N36746          231 GGTAAGCTGT-GCC-TTCA-GCTGTTGCAGATGAAATGTTCTGTTCACCC
1503017         221 GGTAAGCTGT-GCC-TTCA-GCTGTTGCAGATGAAATGTTCTGTTCACCC
1370618         213 GGTAAGCTGT-GCC-TTCA-GC
712299          196 GGTAAGCTGT-GCC-TTNA-GCTGTTTCAGATGAAAGG
2899383         186 GGTAAGCTGT-GCC-TTCA-GCTGTTGCAGATGAAATGTTCTGTTCACCC
2666235         186 GGTAAGCTGT-GCC-TTCA-GCTGTTGCAGATGAAATGTTCTGTTCACCC
2846187         179 GGTAAGCTGT-GCC-TTCA-GCTGTTGCAGATGAAATGTTCTGTTCACCC
1358133         139 NGTAAACTGT-GNC-NTCA-NCTGTTGCAGATGAAATGTTCNGTTCANCC
1349616         135 GGTAAGCTGT-GCC-TTCA-GCTGTTGCAGATGAAATGTTCTGTTCACCC
2305359         132 GGTAAGCTGT-GCC-TTCA-GCTGTTGCAGATGAAATGTTCTGTTCACCC
2149106         129 GGTAAGCTGT-GCC-TTCA-GCTGTTGCAGATGAAATGTTCTGTTCACCC
1984534         129 GGTAAGCTGT-GCC-TTCA-GCTGTTGCAGATGAAATGTTCTGTTCACCC
353611          120 GGTAAGCTGT-GCC-TTCA-GCTGTTGCAGATGAAATGTTCTGTTCACCC
052714          109 GGTAAGCTGT-GCC-TTCA-GCTGTTGCAGATGAAATGTNCTGTTCACCC
AA041443        109 GGTAAGCTGT-GCC-TTCA-GCTGTTGCAGATGAAATGTTCTGTTCACCC
AA090223         93 GGTAAGCTGT-GCC-TTCA-GCTGTTGCAGATGAAATGTTCTGTTCACCC
2152542          93 GGTAAGCTGT-GCC-TTCA-GCTGTTGCAGATGAAATGTTCTGTTCACCC
N26110.RC        89 GGTAAGCTGT-GCC-TTCA-GCTGTTGCAGATGAAATGTTCTGTTCACCC
```

FIGURE 4Z

```
857436          86 GGTAAGCTGT-GCC-TTCA-GCTGTTGCAGATGAAATGTTCTGTTCACCC
1377826         85 GGTAAGCTGT-GCC-TTCA-GCTGTTGCAGATGAAATGTTCTGTTCACCC
                   .++++.++++ +.+ .+.+ .+++++......+++.........+......
<consen02>    1556 GGTAAGCTGT GCC TTCA GCTGTTGCAGATGAAATGTTCTGTTCACCC

H18569          496 CAACCNTGCCATTACCAGG-TTT
HSC3MH091       335 TGCATTACATGTGTTTATT-CA
T78829          301 TGCATTACATGTGTTTATT-CATCCAGCAGTGTTGCTNAGCTCCTACCTC
HSC37E101       297 TGCATTACATGTGTTTATT-CATCCAGCA
N36746          278 TGCATTACATGTGTTTATT-CATCCAGCAGTGTTGCTCAGCTCCTAACTC
1503017         268 TGCAT
2899383         233 TGCATTACATGTGTTTATT-CATCCAG
2666235         233 TGCATTACATGTGTTTAT
2846187         226 TGCATTACATGTGTTTATT-CATCCAGCA
1358133         186 NGNATTACANGTGNTTATT-CATCCAGCAGTGTTGCTCAGCTCCTACCTC
1349616         182 TGCATTACATGTGTTTATT-CATCCAGCAGTGTTGCTCAGCTCCTACCTC
2305359         179 TGCATTACATGTGTTTATT-CATCCAGCAGTGTTGCTCAGCTCCTACCTC
2149106         176 TGCATTACATGTGTTTATT-CATCCAGCAGTGTTGCTCAGCTCCTACCTC
1984534         176 TGCATTACATGTGTTTATT-CATCCAGCAGTGTTGCTCAGCTCCTACCTC
353611          167 TGCATTACATGTGTTTATT-CATCCAGCAGTGTTGCTCAGCTCCTACCTC
052714          156 TGCATTACATGTNTTTATT-CATCCAGCAGTGTTGCTCAGCTCCTACCTC
AA041443        156 TGCATTACATGTGTTTATT-CATCCAGCAGTGTTGCTCAGCTCCTACCTC
AA090223        140 TGCATTACATGTGTTTATT-CATCCAGCAGTGTTGCTCAGCTCCTACCTC
2152542         140 TGCATTACATGTGTNTATT-CATCCAGCAGTGTTGCNCAGCTCCTACCNC
N26110.RC       136 TGCATTACATGTGTTTATTTCATCCAGCAGTGTTGCTCAGCTCCTACCTC
857436          133 TGCATTACATGTGTTTATT-CATCCAGCAGTGTTGCTCAGCTCCTACCTC
1377826         132 TGCATTACATGTGTTTATT-CATCCAGCAGTGTTGCTCAGCTCCTACCTC
                   ..........+....+.. ..++++++++++++..++++++++.+.+
<consen02>    1603 TGCATTACATGTGTTTATT CATCCAGCAGTGTTGCTCAGCTCCTACCTC
```

FIGURE 4AA

```
T78829          350 TGTGCCAGGGGCAGCCATTTTCATATCCCAAGATCAATTCCCTC-TTTTC
N36746          327 TGTGNCAGGG-CAGC-ATTTTCATATCC-AAGATCAATTCCCCCCTCT-C
1358133         235 TGTGCCAGGG-CAGC-ATTTTCATAT
1349616         231 TGTGCCAGGG-CAGC-ATTTTCTT
2305359         228 TGTGCCAGGG-CAGC-ATTTTCATATCC-AAGAT
2149106         225 TG
1984534         225 TGTGCCAGGG-CAGC-ATTTTCATATCC-NAGATCAATNCC
353611          216 TGTGCCAGGG-CAGC-ATTTTCATATCC-AAGATCAT
052714          205 TNTNCCAGGG-CAGC-ATTTTNATATCC-AAGATC
AA041443        205 TGTGCCAGGG-CAGC-ATTTTCATATCC-AAGATCAATTCCCTC-TCT-C
AA090223        189 TGTGCCAGG--CAGC-ATTTTAATATCC-AGGATCAATTCCCTC-TCT-G
2152542         189 TGTGCCAGGG-CNNN-ATTTTCATATCC-AAGATCAATTCCCTC-TCT-C
N26110.RC       186 TGTGCCAGGG-CAGC-ATTTTCATATCC-AAGATCAATTCCCTC-TCT-C
857436          182 TGTGCCAGGG-CAGC-ATTTTCATATCC-AAGATCAATTCCCTC-TNT-C
1377826         181 TGTGCCAGGG-CAGC-ATTTTCATATCC-AAGATCAATTCCCTC-TCT-C
                    +.+..+++++ +... +++++..+++++ ..+++++.+.+++.+ +.+ .
<consen02>     1652 TGTGCCAGGG CAGC ATTTTCATATCC AAGATCAATTCCCTC TCT C

T78829          399 AGGCACAGCCNGGGGGAGGGGGGTCATTG-TTCTTCCTCG-TCCATCAGG
N36746          373 AG-CACAGCCTGGGGGAGGGGGGTCATTGGTTCT-CCTCG-GCCAACAGG
AA041443        250 AG-CACAGCCTGGGG-AGGGGG-TCATTG-TTCT-CCTCGNTCCATCACG
AA090223        233 AG-CACA
2152542         234 AG-CACAGCCTGGGG-AGGG
N26110.RC       231 AG-CACAGCCTGGGG-AGGGGG-TCATTG-TTCT-CCTCG-TCCATCAGG
857436          227 AG-CACAGC-TGGGG-G
1377826         226 AG-CACAGCCTGGG
                    ++ +++++++.++++ .+++++ ++++++ ++++ +++++ .+++.++.+
<consen02>     1697 AG CACAGCCTGGGG AGGGGG TCATTG TTCT CCTCG TCCATCAGG
```

FIGURE 4AB

```
T78829          447 G-GTTNTCAGGAGGGTCAGA
N36746          420 GGAT-CTCAG-AGGCTCAGAGACTGCAAGN-ACTTAACATCTCTGTGCTT
AA041443        295 G-G
N26110.RC       275 G-AT-CTCAG-AGGCTCAGAGACTGCAAGTCACTTAACCTCTCTGTGCTT
                    + .+ .++++ +++.++++++++++++++.++++++++.++++++++++
<consen02>     1741 G AT CTCAG AGGCTCAGAGACTGCAAGTCACTTAACCTCTCTGTGCTT

N36746          467 CAGTCTCCCTGTCNTGTAAAATGGG
N26110.RC       322 CAGTCTCCCTGTCTTGTAAAATGGGAGTAATACCTACCTCACAGGGTTGT
                    ++++++++++++.++++++++++++++++++++++++++++++++++++++
<consen02>     1788 CAGTCTCCCTGTCTTGTAAAATGGGAGTAATACCTACCTCACAGGGTTGT

N26110.RC       372 TGTGGGGATTAATTAGAGATAATGTCTGTAAAGCATTTAAGGTTCTTGAA
                    ++++++++++++++++++++++++++++++++++++++++++++++++++++
<consen02>     1838 TGTGGGGATTAATTAGAGATAATGTCTGTAAAGCATTTAAGGTTCTTGAA

N26110.RC       422 GAAGGCACTATATAAATACAAAATAATATCTATTAAAGTTGGTTTATTTG
                    ++++++++++++++++++++++++++++++++++++++++++++++++++++
<consen02>     1888 GAAGGCACTATATAAATACAAAATAATATCTATTAAAGTTGGTTTATTTG

N26110.RC       472 TG
                    ++
<consen02>     1938 TG
```

FIGURE 5

```
ITB6_CAVPO  Integrin beta-6 subunit - cavia porcellus, fragment (577 aa)
Score = 113 (39.8 bits), Expect = 0.038, P = 0.037
Identities = 38/119 (31%), Positives = 53/119 (44%), at 612,393, Frame = +3

    DNA38268  612 GPSMYCQFASFQYTCQPCRGQRMLCTRDSEC-CGDQLCVWG------HCTKMATR--GSN
                  **  ***  .*  .*  .*  .**  . .* **. .*  *     .**           .
  ITB6_CAVPO  393 GP--YCQCDNF--SCVRHKG--LLCGDNGDCECGECVCRSGWTGEYCNCTTSTDTCISED

    DNA38268  765 GTICDNQRDCQPGLCCAFQRGLLFPVCTPLPVEGELCHDPASRLLDLITWELEPDG----
                  **.*  . **  * *     *   * *  *     * ** .     *   *  **
  ITB6_CAVPO  447 GTLCSGRGDCVCGKCVCTNPGASGPTCERCPT----CSDPCNSKRSCIECHLSADGQPGE

    DNA38268  933 -ALDRCPCA
                    .*.*  *
  ITB6_CAVPO  503 ECVDKCKLA
```

FIGURE 6

```
LFE4_CHICK  Unknown lens fiber protein clfest4 precursor - gallus g (350 aa)
Score = 1003 (353.1 bits), Expect = 2.1e-100, P = 2.1e-100
Identities = 180/293 (61%), Positives = 222/293 (75%), at 285,37, Frame = +3

   DNA38268   285 ATLNEMFREVEELMEDTQHKLRSAVEEMEAEEAAAKASSEVNLANLPPSYHNETNTDTKV
                  *.* ** **** **********.**.****** ** **** ****.****.**.*..
  LFE4_CHICK    37 ASLGEMLREVEALMEDTQHKLRNAVQEMEAEEEGAKKLSEVNFENLPPTYHNESNTETRI

   DNA38268   465 GNNTIHVHREIHKITNNQTGQMVFSETVITSVGDEEGRRSHECIIDEDCGPSMYCQFASF
                  ** *.  *.** *.*.*.** .****.***.  * .*.*********    ****..*
  LFE4_CHICK    97 GNKTVQTHQEIDKVTDNRTGSTIFSETIITSIKGGENKRNHECIIDEDCETGKYCQFSTF

   DNA38268   645 QYTCQPCRGQRMLCTRDSECCGDQLCVWGHCTKMATRGSNGTICDNQRDCQPGLCCAFQR
                  .* ****. *   *.** ********** * *  .** *****.** ** ** *****.
  LFE4_CHICK   157 EYKCQPCKTQHTHCSRDVECCGDQLCVWGECRKATSRGENGTICENQHDCNPGTCCAFQK

   DNA38268   825 GLLFPVCTPLPVEGELCHDPASRLLDLITWELEPDGALDRCPCASGLLCQPHS-HSLVYV
                  ********** *** ****..***.********** *.******** *** * **   *
  LFE4_CHICK   217 ELLFPVCTPLPEEGEPCHDPSNRLLNLITWELEPDGVLERCPCASGLICQPQSSHSTTSV

   DNA38268  1002 CKPTFVGSRDQDGE----------ILLPREVPDEYEVGSFMEEVRQELEDLE
                  *. .  .* . *            *.**.. .** * ..***.*** **
  LFE4_CHICK   277 CELSSNETRKNEKEDPLNMDEMPFISLIPRDILSDYEESSVIQEVRKELESLE
```

H/PKOV 1

PROVISIONAL APPLICATION COVER SHEET

This is a request for filing a PROVISIONAL APPLICATION under 37 CFR 1.53(b)(2)

"EXPRESS MAIL" MAILING LABEL
Number: EM 334 768 055 US Date of Deposit: October 29, 1997
I hereby certify that this paper or fee is being deposited with the United States Postal Service "Express Mail Post Office to Addressee" service under 37 C.F.R. 1.10 on the date indicated above and is addressed to the Assistant Commissioner for Patents, Washington, D.C. 20231 on October 29, 1997.
Typed or Printed Name: Lisa Lee
Signed:

OIPE IAP39 AUG 14 2006 PATENT & TRADEMARK OFFICE

10/29/97 U.S. PTO

Docket Number:	P-65401/WHD/DAV	Type a plus sign (+) inside this box →	+

INVENTOR(s)/APPLICANT(s)

LAST NAME	FIRST NAME	MIDDLE INITIAL	RESIDENCE (CITY AND EITHER STATE OR FOREIGN COUNTRY)
Wood	William	I.	San Mateo, California
Goddard	Audrey		San Francisco, California
Gurney	Austin	L.	Belmont, California

TITLE OF THE INVENTION (*280 characters max*)

NOVEL POLYPEPTIDES HAVING HOMOLOGY TO CD44 ANTIGEN AND NUCLEIC ACIDS ENCODING THE SAME

CORRESPONDENCE ADDRESS

FLEHR HOHBACH TEST ALBRITTON & HERBERT LLP, 4 Embarcadero Center, Suite 3400, San Francisco

STATE	CA	ZIP CODE	94111	COUNTRY	US

ENCLOSED APPLICATION PARTS (*check all that apply*)

X	Specification	*Number of Pages*	38		Small Entity Statement
X	Drawings	*Number of Sheets*	14		Other (*specify*):

METHOD OF PAYMENT OF FILING FEES FOR THIS PROVISIONAL APPLICATION FOR PATENT (*check one*)

X	A check or money order is enclosed to cover the Provisional filing fees	PROVISIONAL FILING FEE AMOUNT ($)	$150
X	The Commissioner is hereby authorized to charge any deficiencies in filing fees or credit any overpayments to Deposit Account Number: 06-1300 (Order No. P-65401/WHD/DAV)		

The invention was made by an agency of the United States Government or under a contract with an agency of the United States Government.

X No.

☐ Yes, the name of the U.S. Government Agency and the Government contract number are: _____

Respectfully submitted, reg no 39,054

SIGNATURE: _____ for

TYPED or PRINTED NAME Walter H. Dreger, Esq.

Date October 29, 1997

REGISTRATION NO. (*if appropriate*) 24,190

60025145.100297

NOVEL POLYPEPTIDES HAVING HOMOLOGY TO CD44 ANTIGEN AND NUCLEIC ACIDS ENCODING THE SAME

FIELD OF THE INVENTION

The present invention relates generally to the identification and isolation of novel DNA and to the recombinant production of novel polypeptides having homology to CD44 antigen, designated herein as "PRO263" polypeptides.

BACKGROUND OF THE INVENTION

CD44 is a cell surface adhesion molecule involved in cell-cell and cell-matrix interactions. Hyaluronic acid, a component of the extracellular matrix is a major ligand. Other ligands include collagen, fibronectin, laminin, chrondroitin sulfate, mucosal addressin, serglycin and osteoponin. CD44 is also important in regulating cell traffic, lymph node homing, transmission of growth signals, and presentation of chemokines and growth factors to traveling cells. CD44 surface proteins are associated with metastatic tumors and CD44 has been used as a marker for HIV infection. Certain splice variants are associated with metastasis and poor prognosis of cancer patients. Therefore, molecules having homology with CD44 are of particular interest, as their homology indicates that they may have functions related to those functions of CD44. CD44 is further described in U.S. Patent Nos. 5,506,119, 5,504,194 and 5,108,904; Gerberick, et al., Toxicol. Appl. Pharmacol., 146(1):1 (1997); Wittig, et al., Immunol. Letters (Netherlands), 57(1-3):217 (1997); and Oliveira and Odell, Oral Oncol. (England), 33(4):260 (1997).

Extracellular and membrane-bound proteins play important roles in the formation, differentiation and maintenance of multicellular organisms. The fate of many individual cells, e.g., proliferation, migration, differentiation, or interaction with other cells, is typically governed by information received from other cells and/or the immediate environment. This information is often transmitted by secreted polypeptides (for instance, mitogenic factors, survival factors, cytotoxic factors, differentiation factors, neuropeptides, and hormones) which are, in turn, received and interpreted by diverse cell receptors or membrane-bound proteins. These secreted polypeptides or signaling molecules normally pass through the cellular

secretory pathway to reach their site of action in the extracellular environment, usually at a membrane-bound receptor protein.

Secreted proteins have various industrial applications, including use as pharmaceuticals, diagnostics, biosensors and bioreactors. In fact, most protein drugs available at present, such as thrombolytic agents, interferons, interleukins, erythropoietins, colony stimulating factors, and various other cytokines, are secretory proteins. Their receptors, which are membrane-bound proteins, also have potential as therapeutic or diagnostic agents. Receptor immunoadhesins, for instance, can be employed as therapeutic agents to block receptor-ligand interaction. Membrane-bound proteins can also be employed for screening of potential peptide or small molecule inhibitors of the relevant receptor/ligand interaction. Such membrane-bound proteins and cell receptors include, but are not limited to, cytokine receptors, receptor kinases, receptor phosphatases, receptors involved in cell-cell interactions, and cellular adhesin molecules like selectins and integrins. Transduction of signals that regulate cell growth and differentiation is regulated in part by phosphorylation of various cellular proteins. Protein tyrosine kinases, enzymes that catalyze that process, can also act as growth factor receptors. Examples include fibroblast growth factor receptor and nerve growth factor receptor.

Efforts are being undertaken by both industry and academia to identify new, native secreted and membrane-bound receptor proteins, particularly transmembrane proteins with homology to CD44 antigen. Many efforts are focused on the screening of mammalian recombinant DNA libraries to identify the coding sequences for novel secreted and membrane-bound receptor proteins. Examples of screening methods and techniques are described in the literature [see, for example, Klein et al., Proc. Natl. Acad. Sci., 93:7108-7113 (1996); U.S. Patent No. 5,536,637)].

SUMMARY OF THE INVENTION

Applicants have identified a cDNA clone that encodes a novel polypeptide having homology to CD44 antigen, wherein the polypeptide is designated in the present application as "PRO263".

In one embodiment, the invention provides an isolated nucleic acid molecule comprising DNA encoding a PRO263 polypeptide. In one aspect, the isolated nucleic acid comprises DNA encoding the PRO263 polypeptide having amino acid residues 1 to 322 of Figure 2 (SEQ ID NO:3), or is complementary to such encoding nucleic acid sequence, and remains stably bound to it under at least moderate, and optionally, under high stringency conditions.

In another embodiment, the invention provides a vector comprising DNA encoding a PRO263 polypeptide. A host cell comprising such a vector is also provided. By way of example, the host cells may be CHO cells, *E. coli*, or yeast. A process for producing PRO263 polypeptides is further provided and comprises culturing host cells under conditions suitable for expression of PRO263 and recovering PRO263 from the cell culture.

In another embodiment, the invention provides isolated PRO263 polypeptide. In particular, the invention provides isolated native sequence PRO263 polypeptide, which in one embodiment, includes an amino acid sequence comprising residues 1 to 322 of Figure 2 (SEQ ID NO:3). An additional embodiment of the present invention is directed to an isolated extracellular domain of a PRO263 polypeptide.

In another embodiment, the invention provides chimeric molecules comprising a PRO263 polypeptide or extracellular domain thereof fused to a heterologous polypeptide or amino acid sequence. An example of such a chimeric molecule comprises a PRO263 polypeptide fused to an epitope tag sequence or a Fc region of an immunoglobulin.

In another embodiment, the invention provides an antibody which specifically binds to a PRO263 polypeptide or extracellular domain thereof. Optionally, the antibody is a monoclonal antibody.

BRIEF DESCRIPTION OF THE DRAWINGS

Figures 1A and 1B show a nucleotide sequence (SEQ ID NO:1) containing the nucleotide sequence (SEQ ID NO:2) of a native sequence PRO263 cDNA, wherein the nucleotide sequence (SEQ ID NO:1) is a clone designated herein as "UNQ230" and/or "DNA34431-seq min". Also presented is the position of the initiator methionine residue, circled, at positions 1-3 of SEQ ID NO: 2 (positions 160-162 of SEQ ID NO: 1). The putative transmembrane domain of the protein is encoded by nucleotides beginning at nucleotide 709 of SEQ ID NO: 2 (underlined). The stop codon is also circled, immediately after the last nucleotide of SEQ ID NO: 2, at position 966.

Figure 2 shows the amino acid sequence (SEQ ID NO:3) derived from SEQ ID NO: 2.

Figures 3A and 3B show an alignment of nucleotide sequences from a variety of expressed sequence tags (SEQ ID NOS: 5-7) as well as a consensus nucleotide sequence derived therefrom designated "DNA30914-from dna" (SEQ ID NO: 4).

Figure 4 shows a BLAST sequence alignment analysis of a portion of the PRO263 amino acid sequence derived from the DNA34431-seq min molecule with a portion of the

CD44 antigen precursor from rat (SEQ ID NO: 8).

Figure 5 shows a BLAST sequence alignment analysis of a portion of the PRO263 amino acid sequence derived from the DNA34431-seq min molecule with a portion of a metastasis-specific variant of the CD44 glycoprotein from rat (SEQ ID NO: 9).

Figure 6 shows a BLAST sequence alignment analysis of a portion of the PRO263 amino acid sequence derived from the DNA34431-seq min molecule with a portion of the CD44 protein from rat (SEQ ID NO: 10).

Figure 7 shows a BLAST sequence alignment analysis of portions of the PRO263 amino acid sequence derived from the DNA34431-seq min molecule with portions of the CD44 antigen precursor from mouse (SEQ ID NOS: 11 and 12).

Figures 8A and 8B show a BLAST sequence alignment analysis of a portion of the PRO263 amino acid sequence derived from the DNA34431-seq min molecule with a portion of the CD44 antigen precursor from cricetulus griseus (SEQ ID NO: 13).

Figure 9 shows a BLAST sequence alignment analysis of a portion of the PRO263 amino acid sequence derived from the DNA34431-seq min molecule with a portion of the CD44 antigen precursor from papio hamadryas (SEQ ID NO: 14).

Figure 10 shows a BLAST sequence alignment analysis of a portion of the PRO263 amino acid sequence derived from the DNA34431-seq min molecule with a portion of the CD44 antigen precursor from mesocricetus auratus (SEQ ID NO: 15).

Figure 11 shows a BLAST sequence alignment analysis of a portion of the PRO263 amino acid sequence derived from the DNA34431-seq min molecule with a portion of the human cell adhesion molecule CD44 (SEQ ID NO: 16).

Figure 12 shows a BLAST sequence alignment analysis of a portion of the PRO263 amino acid sequence derived from the DNA34431-seq min molecule with a portion of the human lymphocyte surface antigen precursor CDW44 (SEQ ID NO: 17).

Figure 13 shows two forward primers (SEQ ID NOS: 18 and 19), a reverse primer (SEQ ID NO: 20) and another primer (SEQ ID NO: 21).

DETAILED DESCRIPTION OF THE PREFERRED EMBODIMENTS

I. Definitions

The terms "PRO263 polypeptide" and "PRO263" when used herein encompass native sequence PRO263 and PRO263 polypeptide variants (which are further defined herein). The PRO263 polypeptides may be isolated from a variety of sources, such as from human tissue types or from another source, or prepared by recombinant or synthetic methods.

A "native sequence PRO263 polypeptide" comprises a polypeptide having the same amino acid sequence as a PRO263 polypeptide derived from nature. Such native sequence PRO263 polypeptide can be isolated from nature or can be produced by recombinant or synthetic means. The term "native sequence PRO263 polypeptide" specifically encompasses naturally-occurring truncated or secreted forms of a PRO263 polypeptide (*e.g.*, an extracellular domain sequence), naturally-occurring variant forms (*e.g.*, alternatively spliced forms) and naturally-occurring allelic variants of a PRO263 polypeptide. In one embodiment of the invention, the native sequence PRO263 polypeptide is a mature or full-length native sequence PRO263 polypeptide comprising amino acids 1 to 322 of Figure 2 (SEQ ID NO:3). In another embodiment of the invention, the native sequence PRO263 polypeptide is an extracellular domain of the full-length PRO263 protein, wherein the putative transmembrane domain of the full-length PRO263 protein is encoded by nucleotides beginning at nucleotide 709 of SEQ ID NO: 2.

"Percent (%) amino acid sequence identity" with respect to the PRO263 amino acid sequences identified herein is defined as the percentage of amino acid residues in a candidate sequence that are identical with the amino acid residues in a PRO263 polypeptide sequence, after aligning the sequences and introducing gaps, if necessary, to achieve the maximum percent sequence identity, and not considering any conservative substitutions as part of the sequence identity. Alignment for purposes of determining percent amino acid sequence identity can be achieved in various ways that are within the skill in the art, for instance, using publicly available computer software such as ALIGN or Megalign (DNASTAR) software. Those skilled in the art can determine appropriate parameters for measuring alignment, including any algorithms needed to achieve maximal alignment over the full length of the sequences being compared.

"Percent (%) nucleic acid sequence identity" with respect to the UNQ230 (DNA34387-seq min) sequence (SEQ ID NO:1) and the coding region therein, SEQ ID NO: 2, shown in Figures 1A and 1B is defined as the percentage of nucleotides in a candidate sequence that are identical with the nucleotides in the UNQ230 (DNA34387-seq min) sequence (SEQ ID NO:1) or SEQ ID NO:2, respectively, after aligning the sequences and introducing gaps, if necessary, to achieve the maximum percent sequence identity. Alignment for purposes of determining percent nucleic acid sequence identity can be achieved in various ways that are within the skill in the art, for instance, using publicly available computer software such as ALIGN or Megalign (DNASTAR) software. Those skilled in the art can determine appropriate parameters for measuring alignment, including any algorithms needed to achieve maximal

alignment over the full length of the sequences being compared.

"Isolated," when used to describe the various polypeptides disclosed herein, means polypeptide that has been identified and separated and/or recovered from a component of its natural environment. Contaminant components of its natural environment are materials that would typically interfere with diagnostic or therapeutic uses for the polypeptide, and may include enzymes, hormones, and other proteinaceous or non-proteinaceous solutes. In preferred embodiments, the polypeptide will be purified (1) to a degree sufficient to obtain at least 15 residues of N-terminal or internal amino acid sequence by use of a spinning cup sequenator, or (2) to homogeneity by SDS-PAGE under non-reducing or reducing conditions using Coomassie blue or, preferably, silver stain. Isolated polypeptide includes polypeptide *in situ* within recombinant cells, since at least one component of the PRO263 polypeptide natural environment will not be present. Ordinarily, however, isolated polypeptide will be prepared by at least one purification step.

An "isolated" PRO263 polypeptide-encoding nucleic acid molecule is a nucleic acid molecule that is identified and separated from at least one contaminant nucleic acid molecule with which it is ordinarily associated in the natural source of the PRO263 polypeptide-encoding nucleic acid. An isolated PRO263 polypeptide-encoding nucleic acid molecule is other than in the former setting in which it is found in nature. Isolated PRO263 polypeptide-encoding nucleic acid molecules therefore are distinguished from the PRO263 polypeptide-encoding nucleic acid molecule as it exists in natural cells. However, an isolated PRO263 polypeptide-encoding nucleic acid molecule includes PRO263 polypeptide-encoding nucleic acid molecules contained in cells that ordinarily express PRO263 polypeptide where, for example, the nucleic acid molecule is in a chromosomal location different from that of natural cells.

The term "control sequences" refers to DNA sequences necessary for the expression of an operably linked coding sequence in a particular host organism. The control sequences that are suitable for prokaryotes, for example, include a promoter, optionally an operator sequence, and a ribosome binding site. Eukaryotic cells are known to utilize promoters, polyadenylation signals, and enhancers.

Nucleic acid is "operably linked" when it is placed into a functional relationship with another nucleic acid sequence. For example, DNA for a presequence or secretory leader is operably linked to DNA for a polypeptide if it is expressed as a preprotein that participates in the secretion of the polypeptide; a promoter or enhancer is operably linked to a coding sequence if it affects the transcription of the sequence; or a ribosome binding site is operably

linked to a coding sequence if it is positioned so as to facilitate translation. Generally, "operably linked" means that the DNA sequences being linked are contiguous, and, in the case of a secretory leader, contiguous and in reading phase. However, enhancers do not have to be contiguous. Linking is accomplished by ligation at convenient restriction sites. If such sites do not exist, the synthetic oligonucleotide adaptors or linkers are used in accordance with conventional practice.

The term "antibody" is used in the broadest sense and specifically covers single anti-PRO263 polypeptide monoclonal antibodies (including agonist, antagonist, and neutralizing antibodies) and anti-PRO263 antibody compositions with polyepitopic specificity. The term "monoclonal antibody" as used herein refers to an antibody obtained from a population of substantially homogeneous antibodies, *i.e.*, the individual antibodies comprising the population are identical except for possible naturally-occurring mutations that may be present in minor amounts.

"Active" or "activity" for the purposes herein refers to form(s) of PRO263 which retain the biologic and/or immunologic activities of native or naturally-occurring PRO263 polypeptide. Variants of PRO263, have the activity of PRO263. Variants of particular interest are those wherein the molecule is truncated or there is a deletion therein so that there is no transdomain region.

II. Compositions and Methods of the Invention

A. Full-length PRO263 Polypeptide

The present invention provides newly identified and isolated nucleotide sequences encoding polypeptides referred to in the present application as PRO263. In particular, Applicants have identified and isolated cDNA encoding a PRO263 polypeptide, as disclosed in further detail in the Examples below. Using BLAST and FastA sequence alignment computer programs, Applicants found that various portions of the PRO263 polypeptide have significant homology with the CD44 antigen and related proteins as shown in Figures 4-12. Accordingly, it is presently believed that PRO263 polypeptide disclosed in the present application is a newly identified member of the CD44 antigen family and possesses at least one of the properties associated with these antigens, i.e., cancer and HIV marker, cell-cell or cell-matrix interactions, regulating cell traffic, lymph node homing, transmission of growth signals, and presentation of chemokines and growth facors to traveling cells.

In addition to the full-length native sequence PRO263 polypeptide described herein, it is contemplated that PRO263 variants can be prepared. PRO263 variants can be prepared by

introducing appropriate nucleotide changes into the PRO263-encoding DNA, or by synthesis of the desired PRO263 polypeptide. Those skilled in the art will appreciate that amino acid changes may alter post-translational processes of the PRO263 polypeptide, such as changing the number or position of glycosylation sites or altering the membrane anchoring characteristics.

Variations in the native full-length sequence PRO263 or in various domains of the PRO263 polypeptide described herein, can be made, for example, using any of the techniques and guidelines for conservative and non-conservative mutations set forth, for instance, in U.S. Patent No. 5,364,934. Variations may be a substitution, deletion or insertion of one or more codons encoding the PRO263 polypeptide that results in a change in the amino acid sequence of the PRO263 polypeptide as compared with the native sequence PRO263. Optionally the variation is by substitution of at least one amino acid with any other amino acid in one or more of the domains of the PRO263 polypeptide. Guidance in determining which amino acid residue may be inserted, substituted or deleted without adversely affecting the desired activity may be found by comparing the sequence of the PRO263 polypeptide with that of homologous known protein molecules and minimizing the number of amino acid sequence changes made in regions of high homology. Amino acid substitutions can be the result of replacing one amino acid with another amino acid having similar structural and/or chemical properties, such as the replacement of a leucine with a serine, i.e., conservative amino acid replacements. Insertions or deletions may optionally be in the range of 1 to 5 amino acids. The variation allowed may be determined by systematically making insertions, deletions or substitutions of amino acids in the sequence and testing the resulting variants for activity in the *in vitro* assay described in the Examples below.

The variations can be made using methods known in the art such as oligonucleotide-mediated (site-directed) mutagenesis, alanine scanning, and PCR mutagenesis. Site-directed mutagenesis [Carter et al., Nucl. Acids Res., 13:4331 (1986); Zoller et al., Nucl. Acids Res., 10:6487 (1987)], cassette mutagenesis [Wells et al., Gene, 34:315 (1985)], restriction selection mutagenesis [Wells et al., Philos. Trans. R. Soc. London SerA, 317:415 (1986)] or other known techniques can be performed on the cloned DNA to produce the PRO263-encoding variant DNA.

Scanning amino acid analysis can also be employed to identify one or more amino acids along a contiguous sequence. Among the preferred scanning amino acids are relatively small, neutral amino acids. Such amino acids include alanine, glycine, serine, and cysteine. Alanine is typically a preferred scanning amino acid among this group because it eliminates the side-

chain beyond the beta-carbon and is less likely to alter the main-chain conformation of the variant. Alanine is also typically preferred because it is the most common amino acid. Further, it is frequently found in both buried and exposed positions [Creighton, The Proteins, (W.H. Freeman & Co., N.Y.); Chothia, J. Mol. Biol., 150:1 (1976)]. If alanine substitution does not yield adequate amounts of variant, an isoteric amino acid can be used.]

B. Modifications of PRO263

Covalent modifications of PRO263 polypeptides are included within the scope of this invention. One type of covalent modification includes reacting targeted amino acid residues of a PRO263 polypeptide with an organic derivatizing agent that is capable of reacting with selected side chains or the N- or C- terminal residues of a PRO263 polypeptide. Derivatization with bifunctional agents is useful, for instance, for crosslinking PRO263 to a water-insoluble support matrix or surface for use in the method for purifying anti-PRO263 antibodies, and vice-versa. Commonly used crosslinking agents include, *e.g.*, 1,1-bis(diazo-acetyl)-2-phenylethane, glutaraldehyde, N-hydroxysuccinimide esters, for example, esters with 4-azidosalicylic acid, homobifunctional imidoesters, including disuccinimidyl esters such as 3,3'-dithiobis(succinimidylpropionate), bifunctional maleimides such as bis-N-maleimido-1,8-octane and agents such as methyl-3-[(p-azidophenyl)dithio]propioimidate.

Other modifications include deamidation of glutaminyl and asparaginyl residues to the corresponding glutamyl and aspartyl residues, respectively, hydroxylation of proline and lysine, phosphorylation of hydroxyl groups of seryl or threonyl residues, methylation of the α-amino groups of lysine, arginine, and histidine side chains [T.E. Creighton, Proteins: Structure and Molecular Properties, W.H. Freeman & Co., San Francisco, pp. 79-86 (1983)], acetylation of the N-terminal amine, and amidation of any C-terminal carboxyl group.

Another type of covalent modification of the PRO263 polypeptide included within the scope of this invention comprises altering the native glycosylation pattern of the polypeptide. "Altering the native glycosylation pattern" is intended for purposes herein to mean deleting one or more carbohydrate moieties found in native sequence PRO263 polypeptide, and/or adding one or more glycosylation sites that are not present in the native sequence PRO263 polypeptide.

Addition of glycosylation sites to PRO263 polypeptides may be accomplished by altering the amino acid sequence thereof. The alteration may be made, for example, by the addition of, or substitution by, one or more serine or threonine residues to the native sequence PRO263 polypeptide (for O-linked glycosylation sites). The PRO263 amino acid sequence

may optionally be altered through changes at the DNA level, particularly by mutating the DNA encoding the PRO263 polypeptide at preselected bases such that codons are generated that will translate into the desired amino acids.

Another means of increasing the number of carbohydrate moieties on the PRO263 polypeptide is by chemical or enzymatic coupling of glycosides to the polypeptide. Such methods are described in the art, e.g., in WO 87/05330 published 11 September 1987, and in Aplin and Wriston, CRC Crit. Rev. Biochem., pp. 259-306 (1981).

Removal of carbohydrate moieties present on the PRO263 polypeptide may be accomplished chemically or enzymatically or by mutational substitution of codons encoding for amino acid residues that serve as targets for glycosylation. Chemical deglycosylation techniques are known in the art and described, for instance, by Hakimuddin, et al., Arch. Biochem. Biophys., 259:52 (1987) and by Edge et al., Anal. Biochem., 118:131 (1981). Enzymatic cleavage of carbohydrate moieties on polypeptides can be achieved by the use of a variety of endo- and exo-glycosidases as described by Thotakura et al., Meth. Enzymol., 138:350 (1987).

Another type of covalent modification of PRO263 comprises linking the PRO263 polypeptide to one of a variety of nonproteinaceous polymers, *e.g.*, polyethylene glycol, polypropylene glycol, or polyoxyalkylenes, in the manner set forth in U.S. Patent Nos. 4,640,835; 4,496,689; 4,301,144; 4,670,417; 4,791,192 or 4,179,337.

PRO263 polypeptides of the present invention may also be modified in a way to form chimeric molecules comprising a PRO263 polypeptide fused to another, heterologous polypeptide or amino acid sequence. In one embodiment, such a chimeric molecule comprises a fusion of a PRO263 polypeptide with a tag polypeptide which provides an epitope to which an anti-tag antibody can selectively bind. The epitope tag is generally placed at the amino- or carboxyl- terminus of the PRO263 polypeptide. The presence of such epitope-tagged forms of a PRO263 polypeptide can be detected using an antibody against the tag polypeptide. Also, provision of the epitope tag enables the PRO263 polypeptide to be readily purified by affinity purification using an anti-tag antibody or another type of affinity matrix that binds to the epitope tag. In an alternative embodiment, the chimeric molecule may comprise a fusion of a PRO263 polypeptide with an immunoglobulin or a particular region of an immunoglobulin. For a bivalent form of the chimeric molecule, such a fusion could be to the Fc region of an IgG molecule.

Various tag polypeptides and their respective antibodies are well known in the art. Examples include poly-histidine (poly-his) or poly-histidine-glycine (poly-his-gly) tags; the flu

HA tag polypeptide and its antibody 12CA5 [Field et al., Mol. Cell. Biol., 8:2159-2165 (1988)]; the c-myc tag and the 8F9, 3C7, 6E10, G4, B7 and 9E10 antibodies thereto [Evan et al., Molecular and Cellular Biology, 5:3610-3616 (1985)]; and the Herpes Simplex virus glycoprotein D (gD) tag and its antibody [Paborsky et al., Protein Engineering, 3(6):547-553 (1990)]. Other tag polypeptides include the Flag-peptide [Hopp et al., BioTechnology, 6:1204-1210 (1988)]; the KT3 epitope peptide [Martin et al., Science, 255:192-194 (1992)]; an α-tubulin epitope peptide [Skinner et al., J. Biol. Chem., 266:15163-15166 (1991)]; and the T7 gene 10 protein peptide tag [Lutz-Freyermuth et al., Proc. Natl. Acad. Sci. USA, 87:6393-6397 (1990)].

C. Preparation of PRO263

The description below relates primarily to production of PRO263 by culturing cells transformed or transfected with a vector containing at least nucleotides 1-966 of SEQ ID NO:2 of the UNQ230 (DNA34431-seq min) nucleic acid (SEQ ID NO:1). It is, of course, contemplated that alternative methods, which are well known in the art, may be employed to prepare PRO263 polypeptides. For instance, the PRO263 sequence, or portions thereof, may be produced by direct peptide synthesis using solid-phase techniques [see, e.g., Stewart et al., Solid-Phase Peptide Synthesis, W.H. Freeman Co., San Francisco, CA (1969); Merrifield, J. Am. Chem. Soc., 85:2149-2154 (1963)]. *In vitro* protein synthesis may be performed using manual techniques or by automation. Automated synthesis may be accomplished, for instance, using an Applied Biosystems Peptide Synthesizer (Foster City, CA) using manufacturer's instructions. Various portions of PRO263 polypeptides may be chemically synthesized separately and combined using chemical or enzymatic methods to produce a full-length PRO263 polypeptide.

1. Isolation of DNA Encoding PRO263

DNA encoding a PRO263 polypeptide may be obtained from a cDNA library prepared from tissue believed to possess the PRO263 mRNA and to express it at a detectable level. Accordingly, human PRO263-encoding DNA can be conveniently obtained from a cDNA library prepared from human tissue, such as described in the Examples. The PRO263-encoding gene may also be obtained from a genomic library or by oligonucleotide synthesis.

Libraries can be screened with probes (such as antibodies to a PRO263 polypeptide or oligonucleotides of at least about 20-80 bases) designed to identify the gene of interest or the protein encoded by it. Screening the cDNA or genomic library with the selected probe may

be conducted using standard procedures, such as described in Sambrook et al., Molecular Cloning: A Laboratory Manual (New York: Cold Spring Harbor Laboratory Press, 1989). An alternative means to isolate the gene encoding PRO263 is to use PCR methodology [Sambrook et al., supra; Dieffenbach et al., PCR Primer:A Laboratory Manual (Cold Spring Harbor Laboratory Press, 1995)].

The Examples below describe techniques for screening a cDNA library. The oligonucleotide sequences selected as probes should be of sufficient length and sufficiently unambiguous that false positives are minimized. The oligonucleotide is preferably labeled such that it can be detected upon hybridization to DNA in the library being screened. Methods of labeling are well known in the art, and include the use of radiolabels like ^{32}P-labeled ATP, biotinylation or enzyme labeling. Hybridization conditions, including moderate stringency and high stringency, are provided in Sambrook et al., supra.

Sequences identified in such library screening methods can be compared and aligned to other known sequences deposited and available in public databases such as GenBank or other private sequence databases. Sequence identity (at either the amino acid or nucleotide level) within defined regions of the molecule or across the full-length sequence can be determined through sequence alignment using computer software programs such as ALIGN, DNAstar, and INHERIT which employ various algorithms to measure homology.

Nucleic acid having protein coding sequence may be obtained by screening selected cDNA or genomic libraries using the deduced amino acid sequence disclosed herein for the first time, and, if necessary, using conventional primer extension procedures as described in Sambrook et al., supra, to detect precursors and processing intermediates of mRNA that may not have been reverse-transcribed into cDNA.

2. Selection and Transformation of Host Cells

Host cells are transfected or transformed with expression or cloning vectors described herein for PRO263 polypeptide production and cultured in conventional nutrient media modified as appropriate for inducing promoters, selecting transformants, or amplifying the genes encoding the desired sequences. The culture conditions, such as media, temperature, pH and the like, can be selected by the skilled artisan without undue experimentation. In general, principles, protocols, and practical techniques for maximizing the productivity of cell cultures can be found in Mammalian Cell Biotechnology: a Practical Approach, M. Butler, ed. (IRL Press, 1991) and Sambrook et al., supra.

Methods of transfection are known to the ordinarily skilled artisan, for example, $CaPO_4$

and electroporation. Depending on the host cell used, transformation is performed using standard techniques appropriate to such cells. The calcium treatment employing calcium chloride, as described in Sambrook et al., supra, or electroporation is generally used for prokaryotes or other cells that contain substantial cell-wall barriers. Infection with *Agrobacterium tumefaciens* is used for transformation of certain plant cells, as described by Shaw et al., Gene, 23:315 (1983) and WO 89/05859 published 29 June 1989. For mammalian cells without such cell walls, the calcium phosphate precipitation method of Graham and van der Eb, Virology, 52:456-457 (1978) can be employed. General aspects of mammalian cell host system transformations have been described in U.S. Patent No. 4,399,216. Transformations into yeast are typically carried out according to the method of Van Solingen et al., J. Bact., 130:946 (1977) and Hsiao et al., Proc. Natl. Acad. Sci. (USA), 76:3829 (1979). However, other methods for introducing DNA into cells, such as by nuclear microinjection, electroporation, bacterial protoplast fusion with intact cells, or polycations, *e.g.*, polybrene, polyornithine, may also be used. For various techniques for transforming mammalian cells, see Keown et al., Methods in Enzymology, 185:527-537 (1990) and Mansour et al., Nature, 336:348-352 (1988).

Suitable host cells for cloning or expressing the DNA in the vectors herein include prokaryote, yeast, or higher eukaryote cells. Suitable prokaryotes include but are not limited to eubacteria, such as Gram-negative or Gram-positive organisms, for example, Enterobacteriaceae such as *E. coli*. Various *E. coli* strains are publicly available, such as *E. coli* K12 strain MM294 (ATCC 31,446); *E. coli* X1776 (ATCC 31,537); *E. coli* strain W3110 (ATCC 27,325) and K5 772 (ATCC 53,635).

In addition to prokaryotes, eukaryotic microbes such as filamentous fungi or yeast are suitable cloning or expression hosts for PRO263-encoding vectors. *Saccharomyces cerevisiae* is a commonly used lower eukaryotic host microorganism.

Suitable host cells for the expression of glycosylated PRO263 are derived from multicellular organisms. Examples of invertebrate cells include insect cells such as Drosophila S2 and Spodoptera Sf9, as well as plant cells. Examples of useful mammalian host cell lines include Chinese hamster ovary (CHO) and COS cells. More specific examples include monkey kidney CV1 line transformed by SV40 (COS-7, ATCC CRL 1651); human embryonic kidney line (293 or 293 cells subcloned for growth in suspension culture, Graham et al., J. Gen Virol., 36:59 (1977)); Chinese hamster ovary cells/-DHFR (CHO, Urlaub and Chasin, Proc. Natl. Acad. Sci. USA, 77:4216 (1980)); mouse sertoli cells (TM4, Mather, Biol. Reprod., 23:243-251 (1980)); human lung cells (W138, ATCC CCL 75); human liver cells

(Hep G2, HB 8065); and mouse mammary tumor (MMT 060562, ATCC CCL51). The selection of the appropriate host cell is deemed to be within the skill in the art.

3. Selection and Use of a Replicable Vector

The nucleic acid (*e.g.*, cDNA or genomic DNA) encoding the desired PRO263 polypeptide may be inserted into a replicable vector for cloning (amplification of the DNA) or for expression. Various vectors are publicly available. The vector may, for example, be in the form of a plasmid, cosmid, viral particle, or phage. The appropriate nucleic acid sequence may be inserted into the vector by a variety of procedures. In general, DNA is inserted into an appropriate restriction endonuclease site(s) using techniques known in the art. Vector components generally include, but are not limited to, one or more of a signal sequence, an origin of replication, one or more marker genes, an enhancer element, a promoter, and a transcription termination sequence. Construction of suitable vectors containing one or more of these components employs standard ligation techniques which are known to the skilled artisan.

The desired PRO263 polypeptide may be produced recombinantly not only directly, but also as a fusion polypeptide with a heterologous polypeptide, which may be a signal sequence or other polypeptide having a specific cleavage site at the N-terminus of the mature protein or polypeptide. In general, the signal sequence may be a component of the vector, or it may be a part of the PRO263-encoding DNA that is inserted into the vector. The signal sequence may be a prokaryotic signal sequence selected, for example, from the group of the alkaline phosphatase, penicillinase, lpp, or heat-stable enterotoxin II leaders. For yeast secretion the signal sequence may be, *e.g.*, the yeast invertase leader, alpha factor leader (including *Saccharomyces* and *Kluyveromyces* α-factor leaders, the latter described in U.S. Patent No. 5,010,182), or acid phosphatase leader, the *C. albicans* glucoamylase leader (EP 362,179 published 4 April 1990), or the signal described in WO 90/13646 published 15 November 1990. In mammalian cell expression, mammalian signal sequences may be used to direct secretion of the protein, such as signal sequences from secreted polypeptides of the same or related species, as well as viral secretory leaders.

Both expression and cloning vectors contain a nucleic acid sequence that enables the vector to replicate in one or more selected host cells. Such sequences are well known for a variety of bacteria, yeast, and viruses. The origin of replication from the plasmid pBR322 is suitable for most Gram-negative bacteria, the 2μ plasmid origin is suitable for yeast, and various viral origins (SV40, polyoma, adenovirus, VSV or BPV) are useful for cloning vectors

in mammalian cells.

Expression and cloning vectors will typically contain a selection gene, also termed a selectable marker. Typical selection genes encode proteins that (a) confer resistance to antibiotics or other toxins, *e.g.*, ampicillin, neomycin, methotrexate, or tetracycline, (b) complement auxotrophic deficiencies, or (c) supply critical nutrients not available from complex media, *e.g.*, the gene encoding D-alanine racemase for *Bacilli*.

An example of suitable selectable markers for mammalian cells are those that enable the identification of cells competent to take up the PRO263-encoding nucleic acid, such as DHFR or thymidine kinase. An appropriate host cell when wild-type DHFR is employed is the CHO cell line deficient in DHFR activity, prepared and propagated as described by Urlaub et al., Proc. Natl. Acad. Sci. USA, 77:4216 (1980). A suitable selection gene for use in yeast is the *trp*1 gene present in the yeast plasmid YRp7 [Stinchcomb et al., Nature, 282:39 (1979); Kingsman et al., Gene, 7:141 (1979); Tschemper et al., Gene, 10:157 (1980)]. The *trp*1 gene provides a selection marker for a mutant strain of yeast lacking the ability to grow in tryptophan, for example, ATCC No. 44076 or PEP4-1 [Jones, Genetics, 85:12 (1977)].

Expression and cloning vectors usually contain a promoter operably linked to the PRO263-encoding nucleic acid sequence to direct mRNA synthesis. Promoters recognized by a variety of potential host cells are well known. Promoters suitable for use with prokaryotic hosts include the β-lactamase and lactose promoter systems [Chang et al., Nature, 275:615 (1978); Goeddel et al., Nature, 281:544 (1979)], alkaline phosphatase, a tryptophan (trp) promoter system [Goeddel, Nucleic Acids Res., 8:4057 (1980); EP 36,776], and hybrid promoters such as the tac promoter [deBoer et al., Proc. Natl. Acad. Sci. USA, 80:21-25 (1983)]. Promoters for use in bacterial systems also will contain a Shine-Dalgarno (S.D.) sequence operably linked to the DNA encoding the PRO263 polypeptide.

Examples of suitable promoting sequences for use with yeast hosts include the promoters for 3-phosphoglycerate kinase [Hitzeman et al., J. Biol. Chem., 255:2073 (1980)] or other glycolytic enzymes [Hess et al., J. Adv. Enzyme Reg., 7:149 (1968); Holland, Biochemistry, 17:4900 (1978)], such as enolase, glyceraldehyde-3-phosphate dehydrogenase, hexokinase, pyruvate decarboxylase, phosphofructokinase, glucose-6-phosphate isomerase, 3-phosphoglycerate mutase, pyruvate kinase, triosephosphate isomerase, phosphoglucose isomerase, and glucokinase.

Other yeast promoters, which are inducible promoters having the additional advantage of transcription controlled by growth conditions, are the promoter regions for alcohol dehydrogenase 2, isocytochrome C, acid phosphatase, degradative enzymes associated with

nitrogen metabolism, metallothionein, glyceraldehyde-3-phosphate dehydrogenase, and enzymes responsible for maltose and galactose utilization. Suitable vectors and promoters for use in yeast expression are further described in EP 73,657.

PRO263 transcription from vectors in mammalian host cells is controlled, for example, by promoters obtained from the genomes of viruses such as polyoma virus, fowlpox virus (UK 2,211,504 published 5 July 1989), adenovirus (such as Adenovirus 2), bovine papilloma virus, avian sarcoma virus, cytomegalovirus, a retrovirus, hepatitis-B virus and Simian Virus 40 (SV40), from heterologous mammalian promoters, *e.g.*, the actin promoter or an immunoglobulin promoter, and from heat-shock promoters, provided such promoters are compatible with the host cell systems.

Transcription of a DNA encoding a PRO263 polypeptide by higher eukaryotes may be increased by inserting an enhancer sequence into the vector. Enhancers are cis-acting elements of DNA, usually about from 10 to 300 bp, that act on a promoter to increase its transcription. Many enhancer sequences are now known from mammalian genes (globin, elastase, albumin, α-fetoprotein, and insulin). Typically, however, one will use an enhancer from a eukaryotic cell virus. Examples include the SV40 enhancer on the late side of the replication origin (bp 100-270), the cytomegalovirus early promoter enhancer, the polyoma enhancer on the late side of the replication origin, and adenovirus enhancers. The enhancer may be spliced into the vector at a position 5' or 3' to the PRO263 coding sequence, but is preferably located at a site 5' from the promoter.

Expression vectors used in eukaryotic host cells (yeast, fungi, insect, plant, animal, human, or nucleated cells from other multicellular organisms) will also contain sequences necessary for the termination of transcription and for stabilizing the mRNA. Such sequences are commonly available from the 5' and, occasionally 3', untranslated regions of eukaryotic or viral DNAs or cDNAs. These regions contain nucleotide segments transcribed as polyadenylated fragments in the untranslated portion of the mRNA encoding PRO263.

Still other methods, vectors, and host cells suitable for adaptation to the synthesis of PRO263 polypeptides in recombinant vertebrate cell culture are described in Gething et al., Nature, 293:620-625 (1981); Mantei et al., Nature, 281:40-46 (1979); EP 117,060; and EP 117,058.

4. Detecting Gene Amplification/Expression

Gene amplification and/or expression may be measured in a sample directly, for example, by conventional Southern blotting, Northern blotting to quantitate the transcription

of mRNA [Thomas, Proc. Natl. Acad. Sci. USA, 77:5201-5205 (1980)], dot blotting (DNA analysis), or *in situ* hybridization, using an appropriately labeled probe, based on the sequences provided herein. Alternatively, antibodies may be employed that can recognize specific duplexes, including DNA duplexes, RNA duplexes, and DNA-RNA hybrid duplexes or DNA-protein duplexes. The antibodies in turn may be labeled and the assay may be carried out where the duplex is bound to a surface, so that upon the formation of duplex on the surface, the presence of antibody bound to the duplex can be detected.

Gene expression, alternatively, may be measured by immunological methods, such as immunohistochemical staining of cells or tissue sections and assay of cell culture or body fluids, to quantitate directly the expression of gene product. Antibodies useful for immunohistochemical staining and/or assay of sample fluids may be either monoclonal or polyclonal, and may be prepared in any mammal. Conveniently, the antibodies may be prepared against a native sequence PRO263 polypeptide or against a synthetic peptide based on the DNA sequences provided herein or against exogenous sequence fused to PRO263-encoding DNA and encoding a specific antibody epitope.

5. Purification of Polypeptide

Forms of PRO263 may be recovered from culture medium or from host cell lysates. If membrane-bound, it can be released from the membrane using a suitable detergent solution (*e.g.* Triton-X 100) or by enzymatic cleavage. Cells employed in expression of PRO263 polypeptides can be disrupted by various physical or chemical means, such as freeze-thaw cycling, sonication, mechanical disruption, or cell lysing agents.

It may be desired to purify PRO263 from recombinant cell proteins or polypeptides. The following procedures are exemplary of suitable purification procedures: by fractionation on an ion-exchange column; ethanol precipitation; reverse phase HPLC; chromatography on silica or on a cation-exchange resin such as DEAE; chromatofocusing; SDS-PAGE; ammonium sulfate precipitation; gel filtration using, for example, Sephadex G-75; protein A Sepharose columns to remove contaminants such as IgG; and metal chelating columns to bind epitope-tagged forms of the PRO263 polypeptide. Various methods of protein purification may be employed and such methods are known in the art and described for example in Deutscher, Methods in Enzymology, 182 (1990); Scopes, Protein Purification:Principles and Practice, Springer-Verlag, New York (1982). The purification step(s) selected will depend, for example, on the nature of the production process used and the particular PRO263 polypeptide produced.

D. Uses for PRO263

Nucleotide sequences (or their complement) encoding PRO263 polypeptides have various applications in the art of molecular biology, including uses as hybridization probes, in chromosome and gene mapping and in the generation of anti-sense RNA and DNA. PRO263-encoding nucleic acid will also be useful for the preparation of PRO263 polypeptides by the recombinant techniques described herein.

The full-length nucleotide sequence SEQ ID NO:1 or the full-length native sequence PRO263 (SEQ ID NO:2) nucleotide sequence, or portions thereof, may be used as hybridization probes for a cDNA library to isolate the full-length PRO263 gene or to isolate still other genes (for instance, those encoding naturally-occurring variants of PRO263 or PRO263 from other species) which have a desired sequence identity to the PRO263 sequence disclosed in Figures 1A and 1B (SEQ ID NOS: 1or 2). Optionally, the length of the probes will be about 20 to about 50 bases. The hybridization probes may be derived from the UNQ230 (DNA34431-seq min) nucleotide sequence of SEQ ID NO:1 or SEQ ID NO: 2 as shown in Figures 1A and 1B or from genomic sequences including promoters, enhancer elements and introns of native sequence PRO263-encoding DNA. By way of example, a screening method will comprise isolating the coding region of the PRO263 gene using the known DNA sequence to synthesize a selected probe of about 40 bases. Hybridization probes may be labeled by a variety of labels, including radionucleotides such as ^{32}P or ^{35}S, or enzymatic labels such as alkaline phosphatase coupled to the probe via avidin/biotin coupling systems. Labeled probes having a sequence complementary to that of the PRO263 gene of the present invention can be used to screen libraries of human cDNA, genomic DNA or mRNA to determine which members of such libraries the probe hybridizes to. Hybridization techniques are described in further detail in the Examples below.

The probes may also be employed in PCR techniques to generate a pool of sequences for identification of closely related PRO263 sequences.

Nucleotide sequences encoding a PRO263 polypeptide can also be used to construct hybridization probes for mapping the gene which encodes that PRO263 polypeptide and for the genetic analysis of individuals with genetic disorders. The nucleotide sequences provided herein may be mapped to a chromosome and specific regions of a chromosome using known techniques, such as *in situ* hybridization, linkage analysis against known chromosomal markers, and hybridization screening with libraries.

When the coding sequences for PRO263 encode a protein which binds to another protein (example, where the PRO263 polypeptide functions as a receptor), the PRO263

polypeptide can be used in assays to identify the other proteins or molecules involved in the binding interaction. By such methods, inhibitors of the receptor/ligand binding interaction can be identified. Proteins involved in such binding interactions can also be used to screen for peptide or small molecule inhibitors or agonists of the binding interaction. Also, the receptor PRO263 polypeptide can be used to isolate correlative ligand(s). Screening assays can be designed to find lead compounds that mimic the biological activity of a native PRO263 or a receptor for PRO263. Such screening assays will include assays amenable to high-throughput screening of chemical libraries, making them particularly suitable for identifying small molecule drug candidates. Small molecules contemplated include synthetic organic or inorganic compounds. The assays can be performed in a variety of formats, including protein-protein binding assays, biochemical screening assays, immunoassays and cell based assays, which are well characterized in the art.

Nucleic acids which encode PRO263 polypeptide or any of its modified forms can also be used to generate either transgenic animals or "knock out" animals which, in turn, are useful in the development and screening of therapeutically useful reagents. A transgenic animal (e.g., a mouse or rat) is an animal having cells that contain a transgene, which transgene was introduced into the animal or an ancestor of the animal at a prenatal, e.g., an embryonic stage. A transgene is a DNA which is integrated into the genome of a cell from which a transgenic animal develops. In one embodiment, cDNA encoding PRO263 polypeptide can be used to clone genomic DNA encoding PRO263 in accordance with established techniques and the genomic sequences used to generate transgenic animals that contain cells which express DNA encoding PRO263. Methods for generating transgenic animals, particularly animals such as mice or rats, have become conventional in the art and are described, for example, in U.S. Patent Nos. 4,736,866 and 4,870,009. Typically, particular cells would be targeted for PRO263 transgene incorporation with tissue-specific enhancers. Transgenic animals that include a copy of a transgene encoding PRO263 introduced into the germ line of the animal at an embryonic stage can be used to examine the effect of increased expression of DNA encoding PRO263. Such animals can be used as tester animals for reagents thought to confer protection from, for example, pathological conditions associated with its overexpression. In accordance with this facet of the invention, an animal is treated with the reagent and a reduced incidence of the pathological condition, compared to untreated animals bearing the transgene, would indicate a potential therapeutic intervention for the pathological condition.

Alternatively, non-human homologues of PRO263 can be used to construct a PRO263 "knock out" animal which has a defective or altered gene encoding PRO263 as a result of

homologous recombination between the endogenous gene encoding PRO263 and altered genomic DNA encoding PRO263 introduced into an embryonic cell of the animal. For example, cDNA encoding PRO263 can be used to clone genomic DNA encoding PRO263 in accordance with established techniques. A portion of the genomic DNA encoding PRO263 can be deleted or replaced with another gene, such as a gene encoding a selectable marker which can be used to monitor integration. Typically, several kilobases of unaltered flanking DNA (both at the 5' and 3' ends) are included in the vector [see e.g., Thomas and Capecchi, Cell, 51:503 (1987) for a description of homologous recombination vectors]. The vector is introduced into an embryonic stem cell line (e.g., by electroporation) and cells in which the introduced DNA has homologously recombined with the endogenous DNA are selected [see e.g., Li et al., Cell, 69:915 (1992)]. The selected cells are then injected into a blastocyst of an animal (e.g., a mouse or rat) to form aggregation chimeras [see e.g., Bradley, in *Teratocarcinomas and Embryonic Stem Cells: A Practical Approach*, E. J. Robertson, ed. (IRL, Oxford, 1987), pp. 113-152]. A chimeric embryo can then be implanted into a suitable pseudopregnant female foster animal and the embryo brought to term to create a "knock out" animal. Progeny harboring the homologously recombined DNA in their germ cells can be identified by standard techniques and used to breed animals in which all cells of the animal contain the homologously recombined DNA. Knockout animals can be characterized for instance, for their ability to defend against certain pathological conditions and for their development of pathological conditions due to absence of the PRO263 polypeptide.

PRO263 can be used in assays wherein CD44 antigen is generally used to determine PRO263 activity relative to that of CD44. The results can be used accordingly.

E. Anti-PRO263 Antibodies

The present invention further provides anti-PRO263 polypeptide antibodies. Exemplary antibodies include polyclonal, monoclonal, humanized, bispecific, and heteroconjugate antibodies.

1. Polyclonal Antibodies

The anti-PRO263 antibodies of the present invention may comprise polyclonal antibodies. Methods of preparing polyclonal antibodies are known to the skilled artisan. Polyclonal antibodies can be raised in a mammal, for example, by one or more injections of an immunizing agent and, if desired, an adjuvant. Typically, the immunizing agent and/or adjuvant will be injected in the mammal by multiple subcutaneous or intraperitoneal injections.

The immunizing agent may include the PRO263 polypeptide or a fusion protein thereof. It may be useful to conjugate the immunizing agent to a protein known to be immunogenic in the mammal being immunized. Examples of such immunogenic proteins include but are not limited to keyhole limpet hemocyanin, serum albumin, bovine thyroglobulin, and soybean trypsin inhibitor. Examples of adjuvants which may be employed include Freund's complete adjuvant and MPL-TDM adjuvant (monophosphoryl Lipid A, synthetic trehalose dicorynomycolate). The immunization protocol may be selected by one skilled in the art without undue experimentation.

2. Monoclonal Antibodies

The anti-PRO263 antibodies may, alternatively, be monoclonal antibodies. Monoclonal antibodies may be prepared using hybridoma methods, such as those described by Kohler and Milstein, Nature, 256:495 (1975). In a hybridoma method, a mouse, hamster, or other appropriate host animal, is typically immunized with an immunizing agent to elicit lymphocytes that produce or are capable of producing antibodies that will specifically bind to the immunizing agent. Alternatively, the lymphocytes may be immunized *in vitro*.

The immunizing agent will typically include the PRO263 polypeptide or a fusion protein thereof. Generally, either peripheral blood lymphocytes ("PBLs") are used if cells of human origin are desired, or spleen cells or lymph node cells are used if non-human mammalian sources are desired. The lymphocytes are then fused with an immortalized cell line using a suitable fusing agent, such as polyethylene glycol, to form a hybridoma cell [Goding, Monoclonal Antibodies: Principles and Practice, Academic Press, (1986) pp. 59-103]. Immortalized cell lines are usually transformed mammalian cells, particularly myeloma cells of rodent, bovine and human origin. Usually, rat or mouse myeloma cell lines are employed. The hybridoma cells may be cultured in a suitable culture medium that preferably contains one or more substances that inhibit the growth or survival of the unfused, immortalized cells. For example, if the parental cells lack the enzyme hypoxanthine guanine phosphoribosyl transferase (HGPRT or HPRT), the culture medium for the hybridomas typically will include hypoxanthine, aminopterin, and thymidine ("HAT medium"), which substances prevent the growth of HGPRT-deficient cells.

Preferred immortalized cell lines are those that fuse efficiently, support stable high level expression of antibody by the selected antibody-producing cells, and are sensitive to a medium such as HAT medium. More preferred immortalized cell lines are murine myeloma lines, which can be obtained, for instance, from the Salk Institute Cell Distribution Center, San

Diego, California and the American Type Culture Collection, Rockville, Maryland. Human myeloma and mouse-human heteromyeloma cell lines also have been described for the production of human monoclonal antibodies [Kozbor, J. Immunol., 133:3001 (1984); Brodeur et al., Monoclonal Antibody Production Techniques and Applications, Marcel Dekker, Inc., New York, (1987) pp. 51-63].

The culture medium in which the hybridoma cells are cultured can then be assayed for the presence of monoclonal antibodies directed against a PRO263 polypeptide. Preferably, the binding specificity of monoclonal antibodies produced by the hybridoma cells is determined by immunoprecipitation or by an *in vitro* binding assay, such as radioimmunoassay (RIA) or enzyme-linked immunoabsorbent assay (ELISA). Such techniques and assays are known in the art. The binding affinity of the monoclonal antibody can, for example, be determined by the Scatchard analysis of Munson and Pollard, Anal. Biochem., 107:220 (1980).

After the desired hybridoma cells are identified, the clones may be subcloned by limiting dilution procedures and grown by standard methods [Goding, supra]. Suitable culture media for this purpose include, for example, Dulbecco's Modified Eagle's Medium and RPMI-1640 medium. Alternatively, the hybridoma cells may be grown *in vivo* as ascites in a mammal.

The monoclonal antibodies secreted by the subclones may be isolated or purified from the culture medium or ascites fluid by conventional immunoglobulin purification procedures such as, for example, protein A-Sepharose, hydroxylapatite chromatography, gel electrophoresis, dialysis, or affinity chromatography.

The monoclonal antibodies may also be made by recombinant DNA methods, such as those described in U.S. Patent No. 4,816,567. DNA encoding the monoclonal antibodies of the invention can be readily isolated and sequenced using conventional procedures (e.g., by using oligonucleotide probes that are capable of binding specifically to genes encoding the heavy and light chains of murine antibodies). The hybridoma cells of the invention serve as a preferred source of such DNA. Once isolated, the DNA may be placed into expression vectors, which are then transfected into host cells such as simian COS cells, Chinese hamster ovary (CHO) cells, or myeloma cells that do not otherwise produce immunoglobulin protein, to obtain the synthesis of monoclonal antibodies in the recombinant host cells. The DNA also may be modified, for example, by substituting the coding sequence for human heavy and light chain constant domains in place of the homologous murine sequences [U.S. Patent No. 4,816,567; Morrison et al., supra] or by covalently joining to the immunoglobulin coding sequence all or part of the coding sequence for a non-immunoglobulin polypeptide. Such a

non-immunoglobulin polypeptide can be substituted for the constant domains of an antibody of the invention, or can be substituted for the variable domains of one antigen-combining site of an antibody of the invention to create a chimeric bivalent antibody.

The antibodies may be monovalent antibodies. Methods for preparing monovalent antibodies are well known in the art. For example, one method involves recombinant expression of immunoglobulin light chain and modified heavy chain. The heavy chain is truncated generally at any point in the Fc region so as to prevent heavy chain crosslinking. Alternatively, the relevant cysteine residues are substituted with another amino acid residue or are deleted so as to prevent crosslinking.

In vitro methods are also suitable for preparing monovalent antibodies. Digestion of antibodies to produce fragments thereof, particularly, Fab fragments, can be accomplished using routine techniques known in the art.

3. Humanized Antibodies

The anti-PRO263 antibodies of the invention may further comprise humanized antibodies or human antibodies. Humanized forms of non-human (e.g., murine) antibodies are chimeric immunoglobulins, immunoglobulin chains or fragments thereof (such as Fv, Fab, Fab', $F(ab')_2$ or other antigen-binding subsequences of antibodies) which contain minimal sequence derived from non-human immunoglobulin. Humanized antibodies include human immunoglobulins (recipient antibody) in which residues from a complementary determining region (CDR) of the recipient are replaced by residues from a CDR of a non-human species (donor antibody) such as mouse, rat or rabbit having the desired specificity, affinity and capacity. In some instances, Fv framework residues of the human immunoglobulin are replaced by corresponding non-human residues. Humanized antibodies may also comprise residues which are found neither in the recipient antibody nor in the imported CDR or framework sequences. In general, the humanized antibody will comprise substantially all of at least one, and typically two, variable domains, in which all or substantially all of the CDR regions correspond to those of a non-human immunoglobulin and all or substantially all of the FR regions are those of a human immunoglobulin consensus sequence. The humanized antibody optimally also will comprise at least a portion of an immunoglobulin constant region (Fc), typically that of a human immunoglobulin [Jones et al., Nature, 321:522-525 (1986); Riechmann et al., Nature, 332:323-329 (1988); and Presta, Curr. Op. Struct. Biol., 2:593-596 (1992)].

Methods for humanizing non-human antibodies are well known in the art. Generally,

a humanized antibody has one or more amino acid residues introduced into it from a source which is non-human. These non-human amino acid residues are often referred to as "import" residues, which are typically taken from an "import" variable domain. Humanization can be essentially performed following the method of Winter and co-workers [Jones et al., Nature, 321:522-525 (1986); Riechmann et al., Nature, 332:323-327 (1988); Verhoeyen et al., Science, 239:1534-1536 (1988)], by substituting rodent CDRs or CDR sequences for the corresponding sequences of a human antibody. Accordingly, such "humanized" antibodies are chimeric antibodies (U.S. Patent No. 4,816,567), wherein substantially less than an intact human variable domain has been substituted by the corresponding sequence from a non-human species. In practice, humanized antibodies are typically human antibodies in which some CDR residues and possibly some FR residues are substituted by residues from analogous sites in rodent antibodies.

Human antibodies can also be produced using various techniques known in the art, including phage display libraries [Hoogenboom and Winter, J. Mol. Biol., 227:381 (1991); Marks et al., J. Mol. Biol., 222:581 (1991)]. The techniques of Cole et al. and Boerner et al. are also available for the preparation of human monoclonal antibodies (Cole et al., Monoclonal Antibodies and Cancer Therapy, Alan R. Liss, p. 77 (1985) and Boerner et al., J. Immunol., 147(1):86-95 (1991)].

4. Bispecific Antibodies

Bispecific antibodies are monoclonal, preferably human or humanized, antibodies that have binding specificities for at least two different antigens. In the present case, one of the binding specificities is for a PRO263 polypeptide, the other one is for any other antigen, and preferably for a cell-surface protein or receptor or receptor subunit.

Methods for making bispecific antibodies are known in the art. Traditionally, the recombinant production of bispecific antibodies is based on the co-expression of two immunoglobulin heavy-chain/light-chain pairs, where the two heavy chains have different specificities [Milstein and Cuello, Nature, 305:537-539 (1983)]. Because of the random assortment of immunoglobulin heavy and light chains, these hybridomas (quadromas) produce a potential mixture of ten different antibody molecules, of which only one has the correct bispecific structure. The purification of the correct molecule is usually accomplished by affinity chromatography steps. Similar procedures are disclosed in WO 93/08829, published 13 May 1993, and in Traunecker et al., EMBO J., 10:3655-3659 (1991).

Antibody variable domains with the desired binding specificities (antibody-antigen

combining sites) can be fused to immunoglobulin constant domain sequences. The fusion preferably is with an immunoglobulin heavy-chain constant domain, comprising at least part of the hinge, CH2, and CH3 regions. It is preferred to have the first heavy-chain constant region (CH1) containing the site necessary for light-chain binding present in at least one of the fusions. DNAs encoding the immunoglobulin heavy-chain fusions and, if desired, the immunoglobulin light chain, are inserted into separate expression vectors, and are co-transfected into a suitable host organism. For further details of generating bispecific antibodies see, for example, Suresh et al., Methods in Enzymology, 121:210 (1986).

5. Heteroconjugate Antibodies

Heteroconjugate antibodies are also within the scope of the present invention. Heteroconjugate antibodies are composed of two covalently joined antibodies. Such antibodies have, for example, been proposed to target immune system cells to unwanted cells [U.S. Patent No. 4,676,980], and for treatment of HIV infection [WO 91/00360; WO 92/200373; EP 03089]. It is contemplated that the antibodies may be prepared *in vitro* using known methods in synthetic protein chemistry, including those involving cross linking agents. For example, immunotoxins may be constructed using a disulfide exchange reaction or by forming a thioether bond. Examples of suitable reagents for this purpose include iminothiolate and methyl-4-mercaptobutyrimidate and those disclosed, for example, in U.S. Patent No. 4,676,980.

F. Uses for anti-PRO263 Antibodies

The anti-PRO263 antibodies of the present invention have various utilities. For example, anti-PRO263 antibodies may be used in diagnostic assays for PRO263 polypeptides, *e.g.*, detecting expression in specific cells, tissues, or serum. Various diagnostic assay techniques known in the art may be used, such as competitive binding assays, direct or indirect sandwich assays and immunoprecipitation assays conducted in either heterogeneous or homogeneous phases [Zola, Monoclonal Antibodies: A Manual of Techniques, CRC Press, Inc. (1987) pp. 147-158]. The antibodies used in the diagnostic assays can be labeled with a detectable moiety. The detectable moiety should be capable of producing, either directly or indirectly, a detectable signal. For example, the detectable moiety may be a radioisotope, such as ^{3}H, ^{14}C, ^{32}P, ^{35}S, or ^{125}I, a fluorescent or chemiluminescent compound, such as fluorescein isothiocyanate, rhodamine, or luciferin, or an enzyme, such as alkaline phosphatase, beta-galactosidase or horseradish peroxidase. Any method known in the art for conjugating the

antibody to the detectable moiety may be employed, including those methods described by Hunter et al., Nature, 144:945 (1962); David et al., Biochemistry, 13:1014 (1974); Pain et al., J. Immunol. Meth., 40:219 (1981); and Nygren, J. Histochem. and Cytochem., 30:407 (1982).

Anti-PRO263 antibodies also are useful for the affinity purification of PRO263 polypeptides from recombinant cell culture or natural sources. In this process, the antibodies against a PRO263 polypeptide are immobilized on a suitable support, such a Sephadex resin or filter paper, using methods well known in the art. The immobilized antibody then is contacted with a sample containing the PRO263 polypeptide to be purified, and thereafter the support is washed with a suitable solvent that will remove substantially all the material in the sample except the PRO263 polypeptide, which is bound to the immobilized antibody. Finally, the support is washed with another suitable solvent that will release the PRO263 polypeptide from the antibody.

The following examples are offered for illustrative purposes only, and are not intended to limit the scope of the present invention in any way.

All patent and literature references cited in the present specification are hereby incorporated by reference in their entirety.

EXAMPLES

Commercially available reagents referred to in the examples were used according to manufacturer's instructions unless otherwise indicated. The source of those cells identified in the following examples, and throughout the specification, by ATCC accession numbers is the American Type Culture Collection, Rockville, Maryland.

EXAMPLE 1: Isolation of cDNA Clones Encoding Human PRO263

The extracellular domain (ECD) sequences (including the secretion signal, if any) of from about 950 known secreted proteins from the Swiss-Prot public protein database were used to search expressed sequence tag (EST) databases. The EST databases included public EST databases (e.g., GenBank) and a proprietary EST DNA database (LIFESEQ™, Incyte Pharmaceuticals, Palo Alto, CA). The search was performed using the computer program BLAST or BLAST2 (Altshul et al., Methods in Enzymology 266:460-480 (1996)) as a comparison of the ECD protein sequences to a 6 frame translation of the EST sequence. Those comparisons resulting in a BLAST score of 70 (or in some cases 90) or greater that did not encode known proteins were clustered and assembled into consensus DNA sequences with

the program "phrap" (Phil Green, University of Washington, Seattle, Washington; http://bozeman.mbt.washington.edu/phrap.docs/phrap.html).

A consensus DNA sequence was assembled relative to other EST sequences using phrap. This consensus sequence is herein designated DNA30914-from dna (see Figures 3A and 3B).

Based on the DNA30914-from dna consensus sequence, oligonucleotides were synthesized: 1) to identify by PCR a cDNA library that contained the sequence of interest, and 2) for use as probes to isolate a clone of the full-length coding sequence for PRO263. Forward and reverse PCR primers generally range from 20 to 30 nucleotides and are often designed to give a PCR product of about 100-1000 bp in length. The probe sequences are typically 40-55 bp in length. In some cases, additional oligonucleotides are synthesized when the consensus sequence is greater than about 1-1.5kbp. In order to screen several libraries for a full-length clone, DNA from the libraries was screened by PCR amplification, as ber Ausubel et al., Current Protocols in Molecular Biology, with the PCR primer pair. A positive library was then used to isolate clones encoding the gene of interest by the *in vivo* clongin procedure using the probe oligonucleotide and one of the primer pairs.

PCR primers were synthesized to form two forward primers (SEQ ID NOS: 18 and 19) and one reverse primer (SEQ ID NO: 20) as shown in Figure 13. Additionally, a synthetic oligonucleotide hybridization probe was constructed from the consensus DNA30914-from dna sequence which has the sequence of SEQ ID NO: 21 also shown in Figure 13.

In order to screen several libraries for a source of a full-length clone, DNA from the libraries was screened by PCR amplification with the PCR primer pair identified above. A positive library was then used to isolate clones encoding the PRO263 gene using the probe oligonucleotide and one of the PCR primers.

RNA for construction of the cDNA libraries was isolated from human fetal liver tissue. The cDNA libraries used to isolate the cDNA clones were constructed by standard methods using commercially available reagents such as those from Invitrogen, San Diego, CA. The cDNA was primed with oligo dT containing a NotI site, linked with blunt to SalI hemikinased adaptors, cleaved with NotI, sized appropriately by gel electrophoresis, and cloned in a defined orientation into a suitable cloning vector (such as pRKB or pRKD; pRK5B is a precursor of pRK5D that does not contain the SfiI site; see, Holmes et al., Science, 253:1278-1280 (1991)) in the unique XhoI and NotI sites.

DNA sequencing of the clones isolated as described above gave the full-length DNA sequence for PRO263 [herein designated as UNQ230 (DNA34431-seq min)] (SEQ ID NO:1)

and the derived protein sequence for PRO263.

The entire nucleotide sequence of UNQ230 (DNA34431-seq min) is shown in Figures 1A and 1B (SEQ ID NO:1). Clone UNQ230 (DNA34431-seq min) contains a single open reading frame with an apparent translational initiation site at nucleotide positions 1-3 [Kozak et al., supra] of SEQ ID NO: 2 and ending at the stop codon after the nucleotide at position 966 of SEQ ID NO: 2 (Figures 1A and 1B). The predicted polypeptide precursor is 322 amino acids long (Figure 2). Clone UNQ230 (DNA34431-seq min) has been deposited with ATCC and is assigned ATCC deposit no.209399.

Analysis of the amino acid sequence of the full-length PRO263 polypeptide suggests that portions of it possess significant homology to CD44 antigen as shown in Figures 4-12, thereby indicating that PRO263 may be a novel cell surface adhesion molecule.

EXAMPLE 2: Use of PRO263-encoding DNA as a hybridization probe

The following method describes use of a nucleotide sequence encoding PRO263 as a hybridization probe.

DNA comprising the coding sequence of full-length PRO263 (as shown in Figures 1A and 1B, SEQ ID NO: 2 or 1) is employed as a probe to screen for homologous DNAs (such as those encoding naturally-occurring variants of PRO263) in human tissue cDNA libraries or human tissue genomic libraries.

Hybridization and washing of filters containing either library DNAs is performed under the following high stringency conditions. Hybridization of radiolabeled UNQ230 (DNA34431-seq min)-derived probe to the filters is performed in a solution of 50% formamide, 5x SSC, 0.1% SDS, 0.1% sodium pyrophosphate, 50 mM sodium phosphate, pH 6.8, 2x Denhardt's solution, and 10% dextran sulfate at 42°C for 20 hours. Washing of the filters is performed in an aqueous solution of 0.1x SSC and 0.1% SDS at 42°C.

DNAs having a desired sequence identity with the DNA encoding full-length native sequence PRO263 polypeptide can then be identified using standard techniques known in the art.

EXAMPLE 3: Expression of PRO263 Polypeptides in *E. coli*

This example illustrates the preparation of unglycosylated forms of PRO263 polypeptides by recombinant expression in *E. coli*.

The DNA sequence encoding PRO263 (SEQ ID NO: 3) is initially amplified using selected PCR primers. The primers should contain restriction enzyme sites which correspond

to the restriction enzyme sites on the selected expression vector. A variety of expression vectors may be employed. An example of a suitable vector is pBR322 (derived from *E. coli*; see Bolivar et al., Gene, 2:95 (1977)) which contains genes for ampicillin and tetracycline resistance. The vector is digested with restriction enzyme and dephosphorylated. The PCR amplified sequences are then ligated into the vector. The vector will preferably include sequences which encode for an antibiotic resistance gene, a trp promoter, a polyhis leader (including the first six STII codons, polyhis sequence, and enterokinase cleavage site), the PRO263 coding region, lambda transcriptional terminator, and an argU gene.

The ligation mixture is then used to transform a selected *E. coli* strain using the methods described in Sambrook et al., supra. Transformants are identified by their ability to grow on LB plates and antibiotic resistant colonies are then selected. Plasmid DNA can be isolated and confirmed by restriction analysis and DNA sequencing.

Selected clones can be grown overnight in liquid culture medium such as LB broth supplemented with antibiotics. The overnight culture may subsequently be used to inoculate a larger scale culture. The cells are then grown to a desired optical density, during which the expression promoter is turned on.

After culturing the cells for several more hours, the cells can be harvested by centrifugation. The cell pellet obtained by the centrifugation can be solubilized using various agents known in the art, and the solubilized PRO263 polypeptide can then be purified using a metal chelating column under conditions that allow tight binding of the polypeptide.

EXAMPLE 4: Expression of PRO263 Polypeptides in Mammalian Cells

This example illustrates preparation of glycosylated forms of PRO263 polypeptides by recombinant expression in mammalian cells.

The vector, pRK5 (see EP 307,247, published March 15, 1989), is employed as the expression vector. Optionally, the PRO263-encoding DNA is ligated into pRK5 with selected restriction enzymes to allow insertion of the PRO263-encoding DNA using ligation methods such as described in Sambrook et al., supra. The resulting vector is called pRK5-PRO263.

In one embodiment, the selected host cells may be 293 cells. Human 293 cells (ATCC CCL 1573) are grown to confluence in tissue culture plates in medium such as DMEM supplemented with fetal calf serum and optionally, nutrient components and/or antibiotics. About 10 μg pRK5-PRO263 DNA is mixed with about 1 μg DNA encoding the VA RNA gene [Thimmappaya et al., Cell, 31:543 (1982)] and dissolved in 500 μl of 1 mM Tris-HCl, 0.1 mM EDTA, 0.227 M $CaCl_2$. To this mixture is added, dropwise, 500 μl of 50 mM HEPES

(pH 7.35), 280 mM NaCl, 1.5 mM $NaPO_4$, and a precipitate is allowed to form for 10 minutes at 25°C. The precipitate is suspended and added to the 293 cells and allowed to settle for about four hours at 37°C. The culture medium is aspirated off and 2 ml of 20% glycerol in PBS is added for 30 seconds. The 293 cells are then washed with serum free medium, fresh medium is added and the cells are incubated for about 5 days.

Approximately 24 hours after the transfections, the culture medium is removed and replaced with culture medium (alone) or culture medium containing 200 μCi/ml ^{35}S-cysteine and 200 μCi/ml ^{35}S-methionine. After a 12 hour incubation, the conditioned medium is collected, concentrated on a spin filter, and loaded onto a 15% SDS gel. The processed gel may be dried and exposed to film for a selected period of time to reveal the presence of PRO263 polypeptide. The cultures containing transfected cells may undergo further incubation (in serum free medium) and the medium is tested in selected bioassays.

In an alternative technique, PRO263-encoding DNA may be introduced into 293 cells transiently using the dextran sulfate method described by Somparyrac et al., Proc. Natl. Acad. Sci., 12:7575 (1981). 293 cells are grown to maximal density in a spinner flask and 700 μg pRK5-PRO263 DNA is added. The cells are first concentrated from the spinner flask by centrifugation and washed with PBS. The DNA-dextran precipitate is incubated on the cell pellet for four hours. The cells are treated with 20% glycerol for 90 seconds, washed with tissue culture medium, and re-introduced into the spinner flask containing tissue culture medium, 5 μg/ml bovine insulin and 0.1 μg/ml bovine transferrin. After about four days, the conditioned media is centrifuged and filtered to remove cells and debris. The sample containing expressed PRO263 polypeptide can then be concentrated and purified by any selected method, such as dialysis and/or column chromatography.

In another embodiment, PRO263 polypeptide can be expressed in CHO cells. The pRK5-PRO263 vector can be transfected into CHO cells using known reagents such as $CaPO_4$ or DEAE-dextran. As described above, the cell cultures can be incubated, and the medium replaced with culture medium (alone) or medium containing a radiolabel such as ^{35}S-methionine. After determining the presence of PRO263 polypeptide, the culture medium may be replaced with serum free medium. Preferably, the cultures are incubated for about 6 days, and then the conditioned medium is harvested. The medium containing the expressed PRO263 polypeptide can then be concentrated and purified by any selected method.

Epitope-tagged PRO263 polypeptide may also be expressed in host CHO cells. The PRO263-encoding DNA may be subcloned out of the pRK5 vector. The subclone insert can undergo PCR to fuse in frame with a selected epitope tag such as a poly-his tag into a

Baculovirus expression vector. The poly-his tagged PRO263-encoding DNA insert can then be subcloned into a SV40 driven vector containing a selection marker such as DHFR for selection of stable clones. Finally, the CHO cells can be transfected (as described above) with the SV40 driven vector. Labeling may be performed, as described above, to verify expression. The culture medium containing the expressed poly-His tagged PRO263 polypeptide can then be concentrated and purified by any selected method, such as by Ni^{2+}-chelate affinity chromatography.

EXAMPLE 5: Expression of a PRO263 Polypeptide in Yeast

The following method describes recombinant expression of PRO263 polypeptides in yeast.

First, yeast expression vectors are constructed for intracellular production or secretion of PRO263 polypeptide from the ADH2/GAPDH promoter. DNA encoding the PRO263 polypeptide of interest, a selected signal peptide and the promoter is inserted into suitable restriction enzyme sites in the selected plasmid to direct intracellular expression of the PRO263 polypeptide. For secretion, DNA encoding the PRO263 polypeptide can be cloned into the selected plasmid, together with DNA encoding the ADH2/GAPDH promoter, the yeast alpha-factor secretory signal/leader sequence, and linker sequences (if needed) for expression of the PRO263 polypeptide.

Yeast cells, such as yeast strain AB110, can then be transformed with the expression plasmids described above and cultured in selected fermentation media. The transformed yeast supernatants can be analyzed by precipitation with 10% trichloroacetic acid and separation by SDS-PAGE, followed by staining of the gels with Coomassie Blue stain.

Recombinant PRO263 polypeptide can subsequently be isolated and purified by removing the yeast cells from the fermentation medium by centrifugation and then concentrating the medium using selected cartridge filters. The concentrate containing the PRO263 polypeptide may further be purified using selected column chromatography resins.

EXAMPLE 6: Expression of PRO263 Polypeptides in Baculovirus

The following method describes recombinant expression of PRO263 polypeptides in Baculovirus.

The PRO263-encoding DNA is fused upstream of an epitope tag contained within a baculovirus expression vector. Such epitope tags include poly-his tags and immunoglobulin tags (like Fc regions of IgG). A variety of plasmids may be employed, including plasmids

derived from commercially available plasmids such as pVL1393 (Novagen). Briefly, the PRO263-encoding DNA or the desired portion of the PRO263-encoding DNA (such as the sequence encoding the extracellular domain of a transmembrane protein) is amplified by PCR with primers complementary to the 5' and 3' regions. The 5' primer may incorporate flanking (selected) restriction enzyme sites. The product is then digested with those selected restriction enzymes and subcloned into the expression vector.

Recombinant baculovirus is generated by co-transfecting the above plasmid and BaculoGold™ virus DNA (Pharmingen) into *Spodoptera frugiperda* ("Sf9") cells (ATCC CRL 1711) using lipofectin (commercially available from GIBCO-BRL). After 4 to 5 days of incubation at 28°C, the released viruses are harvested and used for further amplifications. Viral infection and protein expression is performed as described by O'Reilley et al., Baculovirus expression vectors: A laboratory Manual, Oxford:Oxford University Press (1994).

Expressed poly-his tagged PRO263 polypeptide can then be purified, for example, by Ni^{2+}-chelate affinity chromatography as follows. Extracts are prepared from recombinant virus-infected Sf9 cells as described by Rupert et al., Nature, 362:175-179 (1993). Briefly, Sf9 cells are washed, resuspended in sonication buffer (25 mL Hepes, pH 7.9; 12.5 mM $MgCl_2$; 0.1 mM EDTA; 10% Glycerol; 0.1% NP-40; 0.4 M KCl), and sonicated twice for 20 seconds on ice. The sonicates are cleared by centrifugation, and the supernatant is diluted 50-fold in loading buffer (50 mM phosphate, 300 mM NaCl, 10% Glycerol, pH 7.8) and filtered through a 0.45 μm filter. A Ni^{2+}-NTA agarose column (commercially available from Qiagen) is prepared with a bed volume of 5 mL, washed with 25 mL of water and equilibrated with 25 mL of loading buffer. The filtered cell extract is loaded onto the column at 0.5 mL per minute. The column is washed to baseline A_{280} with loading buffer, at which point fraction collection is started. Next, the column is washed with a secondary wash buffer (50 mM phosphate; 300 mM NaCl, 10% Glycerol, pH 6.0), which elutes nonspecifically bound protein. After reaching A_{280} baseline again, the column is developed with a 0 to 500 mM Imidazole gradient in the secondary wash buffer. One mL fractions are collected and analyzed by SDS-PAGE and silver staining or western blot with Ni^{2+}-NTA-conjugated to alkaline phosphatase (Qiagen). Fractions containing the eluted His_{10}-tagged PRO263 polypeptide are pooled and dialyzed against loading buffer.

Alternatively, purification of the IgG tagged (or Fc tagged) PRO263 polypeptide can be performed using known chromatography techniques, including for instance, Protein A or protein G column chromatography.

EXAMPLE 7: Preparation of Antibodies that Bind PRO263 Polypeptides

This example illustrates the preparation of monoclonal antibodies which can specifically bind to PRO263 polypeptides.

Techniques for producing the monoclonal antibodies are known in the art and are described, for instance, in Goding, supra. Immunogens that may be employed include purified PRO263 polypeptide, fusion proteins containing a PRO263 polypeptide, and cells expressing recombinant PRO263 polypeptide on the cell surface. Selection of the immunogen can be made by the skilled artisan without undue experimentation.

Mice, such as Balb/c, are immunized with the PRO263 immunogen emulsified in complete Freund's adjuvant and injected subcutaneously or intraperitoneally in an amount from 1-100 micrograms. Alternatively, the immunogen is emulsified in MPL-TDM adjuvant (Ribi Immunochemical Research, Hamilton, MT) and injected into the animal's hind foot pads. The immunized mice are then boosted 10 to 12 days later with additional immunogen emulsified in the selected adjuvant. Thereafter, for several weeks, the mice may also be boosted with additional immunization injections. Serum samples may be periodically obtained from the mice by retro-orbital bleeding for testing in ELISA assays to detect anti-PRO263 polypeptide antibodies.

After a suitable antibody titer has been detected, the animals "positive" for antibodies can be injected with a final intravenous injection of PRO263 polypeptide. Three to four days later, the mice are sacrificed and the spleen cells are harvested. The spleen cells are then fused (using 35% polyethylene glycol) to a selected murine myeloma cell line such as P3X63AgU.1, available from ATCC, No. CRL 1597. The fusions generate hybridoma cells which can then be plated in 96 well tissue culture plates containing HAT (hypoxanthine, aminopterin, and thymidine) medium to inhibit proliferation of non-fused cells, myeloma hybrids, and spleen cell hybrids.

The hybridoma cells will be screened in an ELISA for reactivity against PRO263 polypeptide. Determination of "positive" hybridoma cells secreting the desired monoclonal antibodies against a PRO263 polypeptide is within the skill in the art.

The positive hybridoma cells can be injected intraperitoneally into syngeneic Balb/c mice to produce ascites containing the anti-PRO263 polypeptide monoclonal antibodies. Alternatively, the hybridoma cells can be grown in tissue culture flasks or roller bottles. Purification of the monoclonal antibodies produced in the ascites can be accomplished using ammonium sulfate precipitation, followed by gel exclusion chromatography. Alternatively,

affinity chromatography based upon binding of antibody to protein A or protein G can be employed.

Deposit of Material

The following materials have been deposited with the American Type Culture Collection, 12301 Parklawn Drive, Rockville, MD, USA (ATCC):

Material	ATCC Dep. No.	Deposit Date
Plasmid	209399	October 17, 1997

This deposit was made under the provisions of the Budapest Treaty on the International Recognition of the Deposit of Microorganisms for the Purpose of Patent Procedure and the Regulations thereunder (Budapest Treaty). This assures maintenance of a viable culture of the deposit for 30 years from the date of deposit. The deposit will be made available by ATCC under the terms of the Budapest Treaty, and subject to an agreement between Genentech, Inc. and ATCC, which assures permanent and unrestricted availability of the progeny of the culture of the deposit to the public upon issuance of the pertinent U.S. patent or upon laying open to the public of any U.S. or foreign patent application, whichever comes first, and assures availability of the progeny to one determined by the U.S. Commissioner of Patents and Trademarks to be entitled thereto according to 35 USC §122 and the Commissioner's rules pursuant thereto (including 37 CFR §1.14 with particular reference to 886 OG 638).

The assignee of the present application has agreed that if a culture of the materials on deposit should die or be lost or destroyed when cultivated under suitable conditions, the materials will be promptly replaced on notification with another of the same. Availability of the deposited material is not to be construed as a license to practice the invention in contravention of the rights granted under the authority of any government in accordance with its patent laws.

The foregoing written specification is considered to be sufficient to enable one skilled in the art to practice the invention. The present invention is not to be limited in scope by the construct deposited, since the deposited embodiment is intended as a single illustration of certain aspects of the invention and any constructs that are functionally equivalent are within the scope of this invention. The deposit of material herein does not constitute an admission that the written description herein contained is inadequate to enable the practice of any aspect of the invention, including the best mode thereof, nor is it to be construed as limiting the scope of the claims to the specific illustrations that it represents. Indeed, various modifications of

the invention in addition to those shown and described herein will become apparent to those skilled in the art from the foregoing description and fall within the scope of the appended claims.

WHAT IS CLAIMED IS:

1. Isolated nucleic acid comprising DNA encoding a PRO263 polypeptide having amino acid residues 1 to 322 of Figure 2 (SEQ ID NO:3).

2. The nucleic acid of Claim 1 wherein said DNA comprises the nucleotide sequence of SEQ ID NO: 2 or its complement.

3. The nucleic acid of Claim 1 wherein said DNA comprises the nucleotide sequence of SEQ ID NO: 2 or its complement.

4. An isolated nucleic acid comprising the nucleotide sequence of the full-length coding sequence of clone UNQ230 (DNA34431) deposited under accession number ATCC 209399.

5. An isolated nucleic acid encoding an extracellular domain of a PRO263 polypeptide.

6. A vector comprising the nucleic acid of any one of Claim 1, Claim 4 or Claim 5.

7. The vector of Claim 6 operably linked to control sequences recognized by a host cell transformed with the vector.

8. A host cell comprising the vector of Claim 6.

9. The host cell of Claim 8 wherein said cell is a CHO cell.

10. The host cell of Claim 8 wherein said cell is an *E. coli*.

11. The host cell of Claim 8 wherein said cell is a yeast cell.

12. A process for producing a PRO263 polypeptide comprising culturing the host cell of Claim 8 under conditions suitable for expression of said PRO263 polypeptide and recovering said PRO263 polypeptide from the cell culture.

13. Isolated native sequence PRO263 polypeptide comprising amino acid residues 1 to 322 of Figure 2 (SEQ ID NO:3).

14. An isolated extracellular domain of a PRO263 polypeptide.

15. A chimeric molecule comprising a PRO263 polypeptide fused to a heterologous amino acid sequence.

16. The chimeric molecule of Claim 15 wherein said heterologous amino acid sequence is an epitope tag sequence.

17. The chimeric molecule of Claim 15 wherein said heterologous amino acid sequence is a Fc region of an immunoglobulin.

18. An antibody which specifically binds to a PRO263 polypeptide.

19. The antibody of Claim 18 wherein said antibody is a monoclonal antibody.

Abstract of the Disclosure

The present invention is directed to novel polypeptides having homology to CD44 antigen and to nucleic acid molecules encoding those polypeptides. Also provided herein are vectors and host cells comprising those nucleic acid sequences, chimeric polypeptide molecules comprising the polypeptides of the present invention fused to heterologous polypeptide sequences, antibodies which bind to the polypeptides of the present invention and to methods for producing the polypeptides of the present invention.

P01177

```
><
>< /usr/seqdb2/sst/DNA/Dnaseqs.full/ss.DNA34431 (2554 bases)
< good sequence: 3-2554 (2552 bases)
< insert: 84-2455 (2372 bases), 9 regions found
<    5' pRK5 + 1k: 17-901, 65 matches (97%), 28 consec, 2 gaps
<    3' pRK5 + 1k: 2456-1566, 66 matches (100%), 66 consec
<    pRK5D: 2458-1366, 82 matches (97%), 76 consec, 2 gaps
<    pRK5D: 1-885, 37 matches (90%), 30 consec, 2 gaps
<    pRK5D: 22-914, 30 matches (85%), 16 consec, 1 gap
<    pRK5B: 2458-1357, 82 matches (97%), 76 consec, 2 gaps
<    pRK5B: 1-885, 69 matches (97%), 43 consec, 2 gaps
<    3' cDNA linker: 2456-1566, 26 matches (100%), 26 consec
<    5' cDNA linker: 68-950, 16 matches (100%), 16 consec
><                              [DNA34431],
SEQ ID NO: 1 AGCAGGGAAATCCGGATGTCTCGGTTATGAAGTGGAGCAGTGAGTGTGAGCCTCAACATA
GTTCCAGAACTCTCCATCCGGACTAGTTATTGAGCATCTGCCTCTCATATCACCAGTGGC
CATCTGAGGTGTTTCCCTGGCTCTGAAGGGGTAGGCACG
<MET (trans=1-s,dir=f,res=1)
SEQ ID NO: 2 ATGGCCAGGTGCTTCAGCCTGGTGTTGCTTCTCACTTCCATCTGGACCACGAGGCTCCTG
GTCCAAGGCTCTTTGCGTGCAGAAGAGCTTTCCATCCAGGTGTCATGCAGAATTATGGGG
ATCACCCTTGTGAGCAAAAAGGCGAACCAGCAGCTGAATTTCACAGAAGCTAAGGAGGCC
TGTAGGCTGCTGGGACTAAGTTTGGCCGGCAAGGACCAAGTTGAAACAGCCTTGAAAGCT
AGCTTTGAAACTTGCAGCTATGGCTGGGTTGGAGATGGATTCGTGGTCATCTCTAGGATT
AGCCCAAACCCCAAGTGTGGGAAAAATGGGGTGGGTGTCCTGATTTGGAAGGTTCCAGTG
AGCCGACAGTTTGCAGCCTATTGTTACAACTCATCTGATACTTGGACTAACTCGTGCATT
CCAGAAATTATCACCACCAAAGATCCCATATTCAACACTCAAACTGCAACACAAACAACA
GAATTTATTGTCAGTGACAGTACCTACTCGGTGGCATCCCCTTACTCTACAATACCTGCC
CCTACTACTACTCCTCCTGCTCCAGCTTCCACTTCTATTCCACGGAGAAAAAAATTGATT
TGTGTCACAGAAGTTTTTATGGAAACTAGCACCATGTCTACAGAAACTGAACCATTTGTT
GAAAATAAAGCAGCATTCAAGAATGAAGCTGCTGGGTTTGGAGGTGTCCCCACGGCTCTG
CTAGTGCTTGCTCTCCTCTTCTTTGGTGCTGCAGCTGGTCTTGGATTTTGCTATGTCAAA
AGGTATGTGAAGGCCTTCCCTTTTACAAACAAGAATCAGCAGAAGGAAATGATCGAAACC
AAAGTAGTAAAGGAGGAGAAGGCCAATGATAGCAACCCTAATGAGGAATCAAAGAAAACT
```

FIGURE 1A

GATAAAAACCCAGAAGAGTCCAAGAGTCCAAGCAAAACTACCGTGCGATGCCTGGAAGCT
GAAGTTTAGATGAGACAGAAATGAGGAGACACACCTGAGGCTGGTTTCTTTCATGCTCCT
TACCCTGCCCCAGCTGGGGAAATCAAAAGGGCCAAAGAACCAAAGAAGAAAGTCCACCCT
TGGTTCCTAACTGGAATCAGCTCAGGACTGCCATTGGACTATGGAGTGCACCAAAGAGAA
TGCCCTTCTCCTTATTGTAACCCTGTCTGGATCCTATCCTCCTACCTCCAAAGCTTCCCA
CGGCCTTTCTAGCCTGGCTATGTCCTAATAATATCCCACTGGGAGAAAGGAGTTTTGCAA
AGTGCAAGGACCTAAAACATCTCATCAGTATCCAGTGGTAAAAAGGCCTCCTGGCTGTCT
GAGGCTAGGTGGGTTGAAAGCCAAGGAGTCACTGAGACCAAGGCTTTCTCTACTGATTCC
GCAGCTCAGACCCTTTCTTCAGCTCTGAAAGAGAAACACGTATCCCACCTGACATGTCCT
TCTGAGCCCGGTAAGAGCAAAAGAATGGCAGAAAAGTTTAGCCCCTGAAAGCCATGGAGA
TTCTCATAACTTGAGACCTAATCTCTGTAAAGCTAAAATAAAGAAATAGAACAAGGCTGA
GGATACGACAGTACACTGTCAGCAGGGACTGTAAACACAGACAGGGTCAAAGTGTTTTCT
CTGAACACATTGAGTTGGAATCACTGTTTAGAACACACACACTTACTTTTTCTGGTCTCT
ACCACTGCTGATATTTTCTCTAGGAAATATACTTTTACAAGTAACAAAAATAAAAACTCT
TATAAATTTCTATTTTTATCTGAGTTACAGAAATGATTATTAAGGAAGATTACTCAGTAA
TTTGTTTAAAAAGTAATAAAATTCAACAAACATTTGCTGAATAGCTACTATATGTCAAGT
GCTGTGCAAGGTATNACACTCTGTAATTGAATATTATTCCTCAAAAAATTGCACATAGTA
GAACGCTATCTGGGAAGCTATTTTTTTCAGTTTTGATATTTCTAGCTTATCNACTTCCAA
ACTAATTTTTATTTTTTGCTGAGACTAATCTTATTCATTTTCTCTAATATGGCAACCATTA
TAACCTTAATTTATTATTAACATACCTAAGAAGTACATTGTTACNTNTATATACCAAAGC
ACATTTTAAAAGTGCCATTAACAAATGTATCACTAGCCCTCCTTTTTTCCAACAAGAAGGG
ACTGAGAGATGCAGAAATATTTGTGACAAAAAATTAAAGCATTTAGAAAACTT

FIGURE 1B

></usr/seqdb2/sst/DNA/Dnaseqs.min/ss.DNA34431
><subunit 1 of 1, 322 aa, 0 stop
><MW: 35213, pI: 8.71, NX(S/T): 3
MARCFSLVLLLTSIWTTRLLVQGSLRAEELSIQVSCRIMGITLVSKKANQQLNFTEAKEA
CRLLGLSLAGKDQVETALKASFETCSYGWVGDGFVVISRISPNPKCGKNGVGVLIWKVPV
SRQFAAYCYNSSDTWTNSCIPEIITTKDPIFNTQTATQTTEFIVSDSTYSVASPYSTIPA
PTTTPPAPASTSIPRRKKLICVTEVFMETSTMSTETEPFVENKAAFKNEAAGFGGVPTAL
LVLALLFFGAAAGLGFCYVKRYVKAFPFTNKNQQKEMIETKVVKEEKANDSNPNEESKKT
DKNPEESKSPSKTTVRCLEAEV SEQ ID NO: 3

FIGURE 2

<DNA30914: 3 members (1 incyte, 2 est) 623 bp, 4 gaps, 306 bp orf (+3)

```
SEQ ID NO: 5  AA081350          1 GGCTCCTGGTCCAAGGCTCTTTGCGTGCAGAAGAGCTTTCCATCCAGGTG
                                                                      G
SEQ ID NO: 6  1417608           1
SEQ ID NO: 4  <DNA30914>        1 GGCTCCTGGTCCAAGGCTCTTTGCGTGCAGAAGAGCTTTCCATCCAGGTG

              AA081350         51 TCATGCAGAATTATGGGGATCACCCTTGTGAGCAAAAAGGCGAACCAGCA
              1417608           2 TCATGCAGAATTATGGGG-TCACC-TTGTGAGCAAAAAGGCGAACCAGCA
SEQ ID NO: 7  AA046671          1                 GGATCACCCTTGTGAGCAAAAAGGCGAACCAGCA
              <DNA30914>        51 TCATGCAGAATTATGGGGATCACCCTTGTGAGCAAAAAGGCGAACCAGCA

              AA081350        101 GCTGAATTTCACAGAAGCTAAGGAGGCCTGTAGN-TGCTGGGACTAAGTT
              1417608          52 GCTGAATTTCACAGAAGCTAAGGAGGCCTGTAGGCTGCTGGGACTAAGTT
              AA046671         35 GCTGAATTTCACAGAAGCTAAGGAGGCCTGTAGG-TGCTGGGACTAAGTT
              <DNA30914>      101 GCTGAATTTCACAGAAGCTAAGGAGGCCTGTAGGCTGCTGGGACTAAGTT

              AA081350        151 TGGCCGGCAAGGACCAAGTTGAAACAGCCTTGAAAGCTAGCTTTGAAACT
              1417608         102 TGGCCGGCAAGGACCAAGTTGAAACAGCCTTGAAAGCTAGCTTTGAAACT
              AA046671         85 TGGCCGGCAAGGACCAAGTTGAAACAGCCTTGAAAGCTAGCTTTGAAACT
              <DNA30914>      151 TGGCCGGCAAGGACCAAGTTGAAACAGCCTTGAAAGCTAGCTTTGAAACT

              AA081350        201 TGCAGCTATGGCTGGGTTGGAGATGGATTCGTGGTCATCTCCTAGGATTA
              1417608         152 TGCAGCTATGGCTGGGTTGGAGATGGATTCGTGGTCATCTC-TAGGAT-A
              AA046671        135 TGCAGCTATGGCTGGGTTGGAGATGGATTCGTGGTCATCTC-TAGGATTA
              <DNA30914>      201 TGCAGCTATGGCTGGGTTGGAGATGGATTCGTGGTCATCTC-TAGGATTA

              AA081350        251 GCCCAAACCCCAAAGTGTGGGAAAAAATGGGGTGGGTGTCCTGATTTTGG
              1417608         202 GCCCAAACCCCAA-GTGTGGGAAAAA-TGGGGT
              AA046671        185 GCCCAAACCCCAA-GTGTGGGAAAAA-TGGGGTGGGTGTCCTGATTT-GG
              <DNA30914>      251 GCCCAAACCCCAA-GTGTGGGAAAAA-TGGGGTGGGTGTCCTGATTTTGG

              AA081350        301 AAGGTTCCAAGTGAACCGACAAGTTTTGCAGCCTAATTGTTACAACTCAT
              AA046671        235 AAGGTTCCA-GTGAGCCGACA-GTTT-GCAGCCTA-TTGTTACAACTCAT
              <DNA30914>      301 AAGGTTCCAAGTGAACCGACAAGTTTTGCAGCCTAATTGTTACAACTCAT
```

FIGURE 3A

```
AA081350        351 CTGATACNTGGGACTAACTCGTGCCATTCCA
AA046671        285 CTGATACTTGG-ACTAACTCGTGCATTCCAGAAATTATCACCACCAAAGA
<DNA30914>      351 CTGATACTTGG-ACTAACTCGTGCATTCCAGAAATTATCACCACCAAAGA

AA046671        335 TCCCATATTCAACACTCAAACTGCAACACAAACAACAGAATTTATTGTCA
<DNA30914>      401 TCCCATATTCAACACTCAAACTGCAACACAAACAACAGAATTTATTGTCA

AA046671        385 GTGACAGTACCTACTCGGTGGCATCCCCTTACTCTACAATACCTGCCCCT
<DNA30914>      451 GTGACAGTACCTACTCGGTGGCATCCCCTTACTCTACAATACCTGCCCCT

AA046671        435 ACTACTACTCCTCCTGCTCCAGTTCCACTTCTATTCCACGGAGAAAAAAA
<DNA30914>      501 ACTACTACTCCTCCTGCTCCAGTTCCACTTCTATTCCACGGAGAAAAAAA

AA046671        485 TTGATTTGTGTCACAGAAGTTTTTATGGAAACTAGCACCATGTCTACAGA
<DNA30914>      551 TTGATTTGTGTCACAGAAGTTTTTATGGAAACTAGCACCATGTCTACAGA

AA046671        535 AACTGAACCATTTGTTGAAAATAAGCA
<DNA30914>      601 AACTGAACCATTTGTTGAAAATAAGCA
```

FIGURE 3B

1	CD44_RAT	Cd44 antigen precursor - rattus norvegicus	+1	217	52	32
2	P_R14768	Metastasis-specific variant of CD44 glycopr..	+1	217	52	32
3	RNU96138_1	CD44 protein - Rattus norvegicus	+1	217	52	32
4	CD44_MOUSE	Cd44 antigen precursor - mus musculus	+1	208	52	33
5	CD44_CRIGR	Cd44 antigen precursor - cricetulus griseus	+1	204	57	34
6	CD44_PAPHA	Cd44 antigen precursor - papio hamadryas	+1	202	59	33
7	P_R07355	B7 adhesion receptor - Papio sp.	+1	202	59	33
8	CD44_MESAU	Cd44 antigen precursor - mesocricetus aur...	+1	202	56	33
9	JH0417	cell adhesion molecule CD44 - human	+1	200	54	32
10	A32376	lymphocyte surface antigen precursor CDW44 -	+1	200	54	32

>1 CD44_RAT Cd44 antigen precursor - rattus norvegicus (503 aa)
Score = 217 (76.4 bits), Expect = 1.2e-13, P = 1.2e-13
Identities = 52/163 (31%), Positives = 77/163 (47%), at 202,10, Frame = +1

```
      DNA34431  202 WTTRLLVQGSLRAEELSIQVSCRIMGITLVSKKANQQLNFTEAKEACRLLGLSLAGKDQV
                    *    *.* **  ... . ..**  *.  * *      .. *** . *      .*    *.
SEQ ID NO:8 CD44_RAT   10 WGLLCLLQLSLAQQQIDLNITCRYAGVFHVEKNGRYSISRTEAADLCEAFNTTLPTMAQM

      DNA34431  382 ETALKASFETCSYGWVGDGFVVISRISPNPKCGKNGVGVLIWKVPVSRQFAAYCYNSSDT
                    * **.  **** **.. .* *** ** **  *  *  ** *       .  .  **.*.*
      CD44_RAT   70 ELALRKGFETCRYGFI-EGHVVIPRIHPNAICAANNTGVYILLASNTSHYDTYCFNASAP

      DNA34431  562 WTNSCIPEIITTKDPIFNTQTATQTTEFIVSDST-YSVASPYST
                       *    *.  . *.    *   .  * * **    * *
      CD44_RAT  129 LEEDC-----TSVTDLPNSFDGPVTITIVNRDGTRYSKKGEYRT
```

FIGURE 4

>2 P_R14768 Metastasis-specific variant of CD44 glycoprotein - Rattus (503 aa)
Score = 217 (76.4 bits), Expect = 1.2e-13, P = 1.2e-13
Identities = 52/163 (31%), Positives = 77/163 (47%), at 202,10, Frame = +1

```
DNA34431   202 WTTRLLVQGSLRAEELSIQVSCRIMGITLVSKKANQQLNFTEAKEACRLLGLSLAGKDQV
               *    *.* **  ... . ..**  *.  * *      .. *** . *     .*    *.
P_R14768    10 WGLLCLLQLSLAQQQIDLNITCRYAGVFHVEKNGRYSISRTEAADLCEAFNTTLPTMAQM

DNA34431   382 ETALKASFETCSYGWVGDGFVVISRISPNPKCGKNGVGVLIWKVPVSRQFAAYCYNSSDT
               * **.  **** **.. .* *** ** **  *  *  ** *      .  .  **.*.*
P_R14768    70 ELALRKGFETCRYGFI-EGHVVIPRIHPNAICAANNTGVYILLASNTSHYDTYCFNASAP

DNA34431   562 WTNSCIPEIITTKDPIFNTQTATQTTEFIVSDST-YSVASPYST
               *     *.  . *.     *    . * * **    * *
P_R14768   129 LEEDC-----TSVTDLPNSFDGPVTITIVNRDGTRYSKKGEYRT
```

FIGURE 5

>3 RNU96138_1 CD44 protein - Rattus norvegicus (780 aa)
Score = 217 (76.4 bits), Expect = 2.9e-13, P = 2.9e-13
Identities = 52/163 (31%), Positives = 77/163 (47%), at 202,10, Frame = +1

```
DNA34431   202 WTTRLLVQGSLRAEELSIQVSCRIMGITLVSKKANQQLNFTEAKEACRLLGLSLAGKDQV
               *    *.* **  ... . ..**  *.  * *      .. *** . *     .*    *.
RNU96138_1  10 WGLLCLLQLSLAQQQIDLNITCRYAGVFHVEKNGRYSISRTEAADLCEAFNTTLPTMAQM

DNA34431   382 ETALKASFETCSYGWVGDGFVVISRISPNPKCGKNGVGVLIWKVPVSRQFAAYCYNSSDT
               * **.  **** **.. .* *** ** **  *  *  ** *      .  .  **.*.*
RNU96138_1  70 ELALRKGFETCRYGFI-EGHVVIPRIHPNAICAANNTGVYILLASNTSHYDTYCFNASAP

DNA34431   562 WTNSCIPEIITTKDPIFNTQTATQTTEFIVSDST-YSVASPYST
               *     *.  . *.     *    . * * **    * *
RNU96138_1 129 LEEDC-----TSVTDLPNSFDGPVTITIVNRDGTRYSKKGEYRT
```

FIGURE 6

```
>4 CD44_MOUSE  Cd44 antigen precursor - mus musculus (655 aa)
 Score = 208 (73.2 bits), Expect = 2.2e-12, Sum P(2) = 2.2e-12
 Identities = 52/158 (32%), Positives = 77/158 (48%), at 217,14, Frame = +1

   DNA34431    217 LVQGSLRAEELSIQVSCRIMGITLVSKKANQQLNFTEAKEACRLLGLSLAGKDQVETALK
                       *.* **  ... . *.**  *.  * *     .. *** . *.     .*   **.. **
   CD44_MOUSE   14 LLQLSLAHQQIDLNVTCRYAGVFHVEKNGRYSISRTEAADLCQAFNSTLPTMDQMKLALS

   DNA34431    397 ASFETCSYGWVGDGFVVISRISPNPKCGKNGVGVLIWKVPVSRQFAAYCYNSSDTWTNSC
                     **** **.. .* *** ** **  *  *  ** *    .  .  **.*.*        *
   CD44_MOUSE   74 KGFETCRYGFI-EGNVVIPRIHPNAICAANHTGVYILVTSNTSHYDTYCFNASAPPEEDC

   DNA34431    577 IPEIITTKDPIFNTQTATQTTEFIVSDST-YSVASPYST
                        *.  . *.     *   .   * * **   * *
   CD44_MOUSE  133 -----TSVTDLPNSFDGPVTITIVNRDGTRYSKKGEYRT

 Score = 48 (16.9 bits), Expect = 2.2e-12, Sum P(2) = 2.2e-12
 Identities = 12/39 (30%), Positives = 22/39 (56%), at 1006,604, Frame = +1

   DNA34431   1006 VKEEKANDSNPNEESKKTDKNPEESKSPSKTTVRCLEAE
                   *..  * .. *  * **   .     .* **.*  .*. *.
   CD44_MOUSE  604 VEDRKPSELN-GEASKSQEMVHLVNKEPSETPDQCMTAD

 5 CD44_CRIGR  Cd44 antigen precursor - cricetulus griseus (362 aa)
 Score = 204 (71.8 bits), Expect = 1.2e-12, P = 1.2e-12
 Identities = 57/169 (33%), Positives = 80/169 (47%), at 229,18, Frame = +1

   DNA34431    229 SLRAEELSIQVSCRIMGITLVSKKANQQLNFTEAKEACRLLGLSLAGKDQVETALKASFE
                   **  *.. . ..** *.  * *     .. *** . *.   .*   **.   **   **
   CD44_CRIGR   18 SLAHEQIDLNITCRYAGVFHVEKNGRYSISRTEAADLCQAFNSTLPTMDQMVMALSKGFE

   DNA34431    409 TCSYGWVGDGFVVISRISPNPKCGKNGVGVLIWKVPVSRQFAAYCYNSSDTWTNSCIPEI
                   ** **.. .* *** ** **  *  *  ** *      *  .  **.*.*        *
   CD44_CRIGR   78 TCRYGFI-EGHVVIPRIQPNAICAANHTGVYILTSNTSH-YDTYCFNASAPLEEDC----
```

FIGURE 7

FIGURE 8A

SEQ ID NO: 11 (CD44_MOUSE, 1st alignment)

SEQ ID NO: 12 (CD44_MOUSE, 2nd alignment)

SEQ ID NO: 13 (CD44_CRIGR)

```
DNA34431   589 ITTKDPIFNTQTATQTTEFIVSDST-YSVASPYST----IPAPTTTPPAPASTS
               *.  .  *.      *     * * **   * *    * *  **    .* *
CD44_CRIGR 132 -TSVTDLPNSFEGPVTITIVNRDGTRYSKKGEYRTHQEDIDASNTTDDDVSSGS
```

FIGURE 8B

>6 CD44_PAPHA Cd44 antigen precursor - papio hamadryas (362 aa)
Score = 202 (71.1 bits), Expect = 2.1e-12, P = 2.1e-12
Identities = 59/180 (32%), Positives = 84/180 (46%), at 217,14, Frame = +1

```
DNA34431   217 LVQGSLRAEELSIQVSCRIMGITLVSKKANQQLNFTEAKEACRLLGLSLAGKDQVETALK
               *** **    .. . ..**  **  * *      .. *** . *.    .*    *.* **
EQ ID NO:14 CD44_PAPHA  14 LVQLSLA--QIDLNITCRFEGIYHVEKNGRYSISRTEAADLCKAFNSTLPTMAQMEKALS

DNA34431   397 ASFETCSYGWVGDGFVVISRISPNPKCGKNGVGVLIWKVPVSRQFAAYCYNSSDIWTNSC
                 **** **..  .* *** ** **  *  *  ** *     * *.  **.*.*       *
CD44_PAPHA  72 IGFETCRYGFI-EGHVVIPRIHPNSICAANNTGVYILTSNTS-QYDIYCFNASAPPGEDC

DNA34431   577 --IPEIITTKD-PIFNTQTATQTTEFIVSDSTYSVASPYSTIPAPTTTPPAPASTSIPRR
                . ..      * **  *     *  . *    *   .*    *.  *     . .*  *
CD44_PAPHA 130 TSVTDLPNAFDGPITITIVNRDGTRY-VKKGEYRT-NPEDINPSSPTDDDVSSGSSSERS

DNA34431   748 KKL
                *
CD44_PAPHA 188 STL
```

FIGURE 9

```
 8 CD44_MESAU  Cd44 antigen precursor - mesocricetus auratus (431 aa)
  Score = 202 (71.1 bits), Expect = 3.7e-12, P = 3.7e-12
  Identities = 56/169 (33%), Positives = 80/169 (47%), at 229,18, Frame = +1

     DNA34431  229 SLRAEELSIQVSCRIMGITLVSKKANQQLNFTEAKEACRLLGLSLAGKDQVETALKASFE
                   **  ... . ..**  *.  * *     .. *** . *.    .*   **. **   **
FQIDUB:IS CD44_MESAU   18 SLAQQQIDLNITCRYAGVFHVEKNGRYSISRTEAADLCQAFNSTLPTMDQMVMALSKGFE

     DNA34431  409 TCSYGWVGDGFVVISRISPNPKCGKNGVGVLIWKVPVSRQFAAYCYNSSDTWTNSCIPEI
                   ** **.. .* *** ** **  *  *  ** *     *  .  **.*.*      *
  CD44_MESAU   78 TCRYGFI-EGHVVIPRIQPNAICAANHTGVYILTSNTSH-YDTYCFNASAPLEEDC----

     DNA34431  589 ITTKDPIFNTQTATQTTEFIVSDST-YSVASPYST----IPAPTTTPPAPASTS
                    *.  . *.     *   . * * **    * *    * *  **     .* *
  CD44_MESAU  132 -TSVTDLPNSFEGPVTITIVNRDGTRYSKKGEYRTHQEDIDASNTTDDDVSSGS
```

FIGURE 10

>9 JH0417 cell adhesion molecule CD44 - human (361 aa)
Score = 200 (70.4 bits), Expect = 3.5e-12, P = 3.5e-12
Identities = 54/168 (32%), Positives = 80/168 (47%), at 232,17, Frame = +1

```
DNA34431  232 LRAEELSIQVSCRIMGITLVSKKANQQLNFTEAKEACRLLGLSLAGKDQVETALKASFET
              *  . . . ..**  *.  * *     .. *** . *.    .*    *.* **   ***
JH0417     17 LSLAQIDLNITCRFAGVFHVEKNGRYSISRTEAADLCKAFNSTLPTMAQMEKALSIGFET

DNA34431  412 CSYGWVGDGFVVISRISPNPKCGKNGVGVLIWKVPVSRQFAAYCYNSSDIWTNSC--IPE
              * **.. .* *** ** **  *  *  ** *     * *.  **.*.*       *  . .
JH0417     77 CRYGFI-EGHVVIPRIHPNSICAANNTGVYILTYNTS-QYDTYCFNASAPPEEDCTSVTD

DNA34431  586 IITTKD-PIFNTQTATQTTEFIVSDSTYSVASPYSTIPA-PTTTPPAPASTS
              .     * **  *      * . *    *    .*    *. **    .  *.*
JH0417    135 LPNAFDGPITITIVNRDGTRY-VQKGEYRT-NPEDIYPSNPTDDDVSSGSSS
```

SEQ ID NO: 16

FIGURE 11

>10 A32376 lymphocyte surface antigen precursor CDW44 - human (361 aa)
Score = 200 (70.4 bits), Expect = 3.5e-12, P = 3.5e-12
Identities = 54/168 (32%), Positives = 80/168 (47%), at 232,17, Frame = +1

```
   DNA34431  232 LRAEELSIQVSCRIMGITLVSKKANQQLNFTEAKEACRLLGLSLAGKDQVETALKASFET
                 *   . . . ..**  *.  * *     .. *** . *.    .*    *.* **   ***
SEQ ID NO:17  A32376   17 LSLAQIDLNITCRFAGVFHVEKNGRYSISRTEAADLCKAFNSTLPTMAQMEKALSIGFET

   DNA34431  412 CSYGWVGDGFVVISRISPNPKCGKNGVGVLIWKVPVSRQFAAYCYNSSDTWTNSC--IPE
                 * **.. .* *** ** **  *  *  ** *     * *.  **.*.*      *  . .
     A32376   77 CRYGFI-EGHVVIPRIHPNSICAANNTGVYILTYNTS-QYDTYCFNASAPPEEDCTSVTD

   DNA34431  586 IITTKD-PIFNTQTATQTTEFIVSDSTYSVASPYSTIPA-PTTTPPAPASTS
                 .    * **  *      * . *    *   .*    *. **    .  *.*
     A32376  135 LPNAFDGPITITIVNRDGTRY-VQKGEYRT-NPEDIYPSNPTDDDVSSGSSS
```

FIGURE 12

30914.p

AGGAGGCCTGTAGGCTGCTGGGACTAAGTTTGGCCGGCAAGGACCAAGTT SEQ ID NO: 21

.f1

GAGCTTTCCATCCAGGTGTCATGC SEQ ID NO: 18

.r

TGGAGCAGGAGGAGTAGTAGTAGG SEQ ID NO: 20

,f2

GTCAGTGACAGTACCTACTCGG SEQ ID NO: 19

FIGURE 13

A/PROV

PROVISIONAL APPLICATION COVER SHEET

This is a request for filing a PROVISIONAL APPLICATION under 37 CFR 1.53(b)(2)

"EXPRESS MAIL" MAILING LABEL
Number: EM 104 543 068 US Date of Deposit: April 22, 1998
I hereby certify that this paper or fee is being deposited with the United States Postal Service "Express Mail Post Office to Addressee" service under 37 C.F.R. 1.10 on the date indicated above and is addressed to Box Provisional Patent Application, Assistant Commissioner for Patents, Washington, D.C. 20231, on April 22, 1998.
Typed or Printed Name: Kevin Campbell
Signed: K Campbell

Docket Number:	P-65410/WHD/MTK Genentech Docket No. PR1189	Type a plus sign (+) inside this box →	+

INVENTOR(s)/APPLICANT(s)

LAST NAME	FIRST NAME	MIDDLE INITIAL	RESIDENCE (CITY AND EITHER STATE OR FOREIGN COUNTRY)
Wood	William		San Mateo, California
Goddard	Audrey		San Francisco, California
Gurney	Austin		Belmont, California

TITLE OF THE INVENTION (*280 characters max*)

NOVEL POLYPEPTIDES HAVING HOMOLOGY TO DIFF33 AND NUCLEIC ACIDS ENCODING THEREFOR

CORRESPONDENCE ADDRESS

Walter H. Dreger, Registration No. 24, 190
FLEHR HOHBACH TEST ALBRITTON & HERBERT LLP, 4 Embarcadero Center, Suite 3400, San Francisco

STATE	CA	ZIP CODE	94111	COUNTRY	US

ENCLOSED APPLICATION PARTS (*check all that apply*)

X	Specification	*Number of Pages*	38		Small Entity Statement
X	Drawings	*Number of Sheets*	38		Other (*specify*):

METHOD OF PAYMENT OF FILING FEES FOR THIS PROVISIONAL APPLICATION FOR PATENT (*check one*)

X	A check or money order is enclosed to cover the Provisional filing fees	PROVISIONAL FILING FEE AMOUNT ($) $150/$75	$150
	The Commissioner is hereby authorized to charge filing fees and credit Deposit Account Number: 06-1300 (Order No. P-65410/WHD/MTK)		

The invention was made by an agency of the United States Government or under a contract with an agency of the United States Government.

X No.

☐ Yes, the name of the U.S. Government Agency and the Government contract number are: _____

Respectfully submitted,

SIGNATURE: Mark T. Kresnak

Date: April 22, 1998

TYPED or PRINTED NAME: Mark T. Kresnak, Ph.D.

REGISTRATION NO. (*if appropriate*): P-42,767

☐ Additional inventors are being named on separately numbered sheets attached hereto.

USE ONLY FOR FILING A PROVISIONAL APPLICATION FOR PATENT

Burden Hour Statement: This form is estimated to take .2 hours to complete. Times will vary depending upon the needs of the individual case. Any comments on the amount of time you are required to complete this form should be sent to the Chief Information Officer, Assistant Commissioner for Patents, Washington, DC 20231. DO NOT SEND FEES OR COMPLETED FORMS TO THIS ADDRESS. SEND TO: Box Provisional Application, Assistant Commissioner for Patents, Washington, D.C. 20231.

585617

NOVEL POLYPEPTIDES HAVING HOMOLOGY TO DIFF33 AND NUCLEIC ACIDS ENCODING THEREFOR

FIELD OF THE INVENTION

The present invention relates generally to the identification and isolation of novel DNA and to the recombinant production of novel polypeptides having homology to the human Diff 33 protein, designated herein as "PRO300" polypeptides.

BACKGROUND OF THE INVENTION

Membrane-bound proteins and receptors can play an important role in the formation, differentiation and maintenance of multicellular organisms. The fate of many individual cells, e.g., proliferation, migration, differentiation, or interaction with other cells, is typically governed by information received from other cells and/or the immediate environment. This information is often transmitted by secreted polypeptides (for instance, mitogenic factors, survival factors, cytotoxic factors, differentiation factors, neuropeptides, and hormones) which are, in turn, received and interpreted by diverse cell receptors or membrane-bound proteins. Such membrane-bound proteins and cell receptors include, but are not limited to, cytokine receptors, receptor kinases, receptor phosphatases, receptors involved in cell-cell interactions, and cellular adhesin molecules like selectins and integrins. For instance, transduction of signals that regulate cell growth and differentiation is regulated in part by phosphorylation of various cellular proteins. Protein tyrosine kinases, enzymes that catalyze that process, can also act as growth factor receptors. Examples include fibroblast growth factor receptor and nerve growth factor receptor.

Membrane-bound proteins and receptor molecules have various industrial applications, including as pharmaceutical and diagnostic agents. Receptor immunoadhesins, for instance, can be employed as therapeutic agents to block receptor-ligand interaction. The membrane-bound proteins can also be employed for screening of potential peptide or small molecule inhibitors of the relevant receptor/ligand interaction.

Efforts are being undertaken by both industry and academia to identify new, native receptor proteins. Many efforts are focused on the screening of mammalian recombinant DNA libraries to identify the coding sequences for novel receptor proteins.

The Diff 33 protein is over-expressed in mouse testicular tumors. At present its role is unclear, however, it may play a role in cancer. Given the medical importance of understanding the physiology of cancer, efforts are currently being under taken to identify new, native proteins which are involved in cancer. We describe herein the identification of a novel polypeptide which has homology to Diff 33.

SUMMARY OF THE INVENTION

Applicants have identified a cDNA clone that encodes a novel polypeptide, wherein the polypeptide is designated in the present application as "PRO300".

In one embodiment, the invention provides an isolated nucleic acid molecule comprising DNA encoding a PRO300 polypeptide. In one aspect, the isolated nucleic acid comprises DNA encoding the PRO300 polypeptide having amino acid residues 1 to 457 of Figure 2 (SEQ ID NO:3), or is complementary to such encoding nucleic acid sequence, and remains stably bound to it under at least moderate, and optionally, under high stringency conditions. The isolated nucleic acid sequence may comprise the cDNA insert of the DNA40625-1189 vector deposited on April 21, 1998 as ATCC _____ which includes the nucleotide sequence encoding PRO300.

In another embodiment, the invention provides a vector comprising DNA encoding a PRO300 polypeptide. A host cell comprising such a vector is also provided. By way of example, the host cells may be CHO cells, *E. coli*, or yeast. A process for producing PRO300 polypeptides is further provided and comprises culturing host cells under conditions suitable for expression of PRO300 and recovering PRO300 from the cell culture.

In another embodiment, the invention provides isolated PRO300 polypeptide. In particular, the invention provides isolated native sequence PRO300 polypeptide, which in one embodiment, includes an amino acid sequence comprising residues 1 to 457 of Figure 2 (SEQ ID NO:3). An additional embodiment of the present invention is directed to an isolated extracellular domain of a PRO300 polypeptide. Optionally, the PRO300 polypeptide is obtained or is obtainable by expressing the polypeptide encoded by the cDNA insert of the DNA40625-1189 vector deposited on April 21, 1998 as ATCC _____.

In another embodiment, the invention provides chimeric molecules comprising a PRO300 polypeptide or extracellular domain thereof fused to a heterologous polypeptide or amino acid sequence. An example of such a chimeric molecule comprises a PRO300 polypeptide fused to an epitope tag sequence or a Fc region of an immunoglobulin.

In another embodiment, the invention provides an antibody which specifically binds to a PRO300 polypeptide or extracellular doain thereof. Optionally, the antibody is a monoclonal antibody.

BRIEF DESCRIPTION OF THE DRAWINGS

Figure 1 shows a nucleotide sequence (SEQ ID NO:1) containing the nucleotide sequence (SEQ ID NO:2) of a native sequence PRO300 cDNA (nucleotides 45-1415), wherein the nucleotide sequence (SEQ ID NO:1) is designated herein as "UNQ263" and/or "DNA40625-1189". Also presented is the position of the initiator methionine residue.

Figure 2 shows the amino acid sequence (SEQ ID NO:3) derived from nucleotides 45-1415 of the nucleotide sequence shown in Figure 1.

Figures 3A-3B show an alignment of portions of the nucleotide sequences from a variety of expressed sequence tags as well as an intermediate consensus nucleotide sequence assembled therefrom designated "DNA35930".

Figures 4A through 4AH show an alignment of portions of the nucleotide sequences from a variety of expressed sequence tags as well as an extended consensus nucleotide sequence assembled therefrom using repeated cycles of BLAST and phrap, wherein the consensus sequence is herein designated <consen01>.

DETAILED DESCRIPTION OF THE PREFERRED EMBODIMENTS

I. Definitions

The terms "PRO300 polypeptide" and "PRO300" when used herein encompass native sequence PRO300 and PRO300 polypeptide variants (which are further defined herein). The PRO300 polypeptides may be isolated from a variety of sources, such as from human tissue types or from another source, or prepared by recombinant or synthetic methods.

A "native sequence PRO300 polypeptide" comprises a polypeptide having the same amino acid sequence as a PRO300 polypeptide derived from nature. Such native sequence PRO300 polypeptide can be isolated from nature or can be produced by recombinant or

synthetic means. The term "native sequence PRO300 polypeptide" specifically encompasses naturally-occurring truncated or secreted forms of a PRO300 polypeptide (*e.g.*, an extracellular domain sequence), naturally-occurring variant forms (*e.g.*, alternatively spliced forms) and naturally-occurring allelic variants of a PRO300 polypeptide. In one embodiment of the invention, the native sequence PRO300 polypeptide is a mature or full-length native sequence PRO300 polypeptide comprising amino acids 1 to 457 of Figure 2 (SEQ ID NO:3). In another embodiment of the invention, the native sequence PRO300 polypeptide is an extracellular domain of the full-length PRO300 protein. Optionally, the PRO300 polypeptide is obtained or obtainable by expressing the polypeptide encoded by the cDNA insert of the vector DNA40625-1189 deposited on April 21, 1998 as ATCC _____.

"Percent (%) amino acid sequence identity" with respect to the PRO300 amino acid sequences identified herein is defined as the percentage of amino acid residues in a candidate sequence that are identical with the amino acid residues in a PRO300 polypeptide sequence, after aligning the sequences and introducing gaps, if necessary, to achieve the maximum percent sequence identity, and not considering any conservative substitutions as part of the sequence identity. Alignment for purposes of determining percent amino acid sequence identity can be achieved in various ways that are within the skill in the art, for instance, using publicly available computer software such as ALIGN or Megalign (DNASTAR) software. Those skilled in the art can determine appropriate parameters for measuring alignment, including any algorithms needed to achieve maximal alignment over the full length of the sequences being compared.

"Percent (%) nucleic acid sequence identity" with respect to the UNQ263 (DNA40625-1189) sequence (SEQ ID NO:1) and nucleotides 45-1415 of the sequence shown in Figure 1 (SEQ ID NO:2) is defined as the percentage of nucleotides in a candidate sequence that are identical with the nucleotides in the UNQ263 (DNA40625-1189) sequence (SEQ ID NO:1) or nucleotides nucleotides 45-1415 of the sequnece shown in Figure 1 (SEQ ID NO:2), respectively, after aligning the sequences and introducing gaps, if necessary, to achieve the maximum percent sequence identity. Alignment for purposes of determining percent nucleic acid sequence identity can be achieved in various ways that are within the skill in the art, for instance, using publicly available computer software such as ALIGN or Megalign (DNASTAR) software. Those skilled in the art can determine appropriate parameters for measuring alignment, including any algorithms needed to achieve maximal alignment over the

full length of the sequences being compared.

The term "epitope tagged" where used herein refers to a chimeric polypeptide comprising a PRO300 polypeptide, or domain sequence thereof, fused to a "tag polypeptide". The tag polypeptide has enough residues to provide an epitope against which an antibody may be made, or which can be identified by some other agent, yet is short enough such that it does not interfere with the activity of the PRO300 polypeptide. The tag polypeptide preferably is also fairly unique so that the antibody does not substantially cross-react with other epitopes. Suitable tag polypeptides generally have at least six amino acid residues and usually between about 8 to about 50 amino acid residues (preferably, between about 10 to about 20 residues).

"Isolated," when used to describe the various polypeptides disclosed herein, means polypeptide that has been identified and separated and/or recovered from a component of its natural environment. Contaminant components of its natural environment are materials that would typically interfere with diagnostic or therapeutic uses for the polypeptide, and may include enzymes, hormones, and other proteinaceous or non-proteinaceous solutes. In preferred embodiments, the polypeptide will be purified (1) to a degree sufficient to obtain at least 15 residues of N-terminal or internal amino acid sequence by use of a spinning cup sequenator, or (2) to homogeneity by SDS-PAGE under non-reducing or reducing conditions using Coomassie blue or, preferably, silver stain. Isolated polypeptide includes polypeptide *in situ* within recombinant cells, since at least one component of the PRO300 polypeptide natural environment will not be present. Ordinarily, however, isolated polypeptide will be prepared by at least one purification step.

An "isolated" PRO300 polypeptide-encoding nucleic acid molecule is a nucleic acid molecule that is identified and separated from at least one contaminant nucleic acid molecule with which it is ordinarily associated in the natural source of the PRO300 polypeptide-encoding nucleic acid. An isolated PRO300 polypeptide-encoding nucleic acid molecule is other than in the former setting in which it is found in nature. Isolated PRO300 polypeptide-encoding nucleic acid molecules therefore are distinguished from the PRO300 polypeptide-encoding nucleic acid molecule as it exists in natural cells. However, an isolated PRO300 polypeptide-encoding nucleic acid molecule includes PRO300 polypeptide-encoding nucleic acid molecules contained in cells that ordinarily express PRO300 polypeptide where, for example, the nucleic acid molecule is in a chromosomal location different from that of natural cells.

The term "control sequences" refers to DNA sequences necessary for the expression of an operably linked coding sequence in a particular host organism. The control sequences that are suitable for prokaryotes, for example, include a promoter, optionally an operator sequence, and a ribosome binding site. Eukaryotic cells are known to utilize promoters, polyadenylation signals, and enhancers.

Nucleic acid is "operably linked" when it is placed into a functional relationship with another nucleic acid sequence. For example, DNA for a presequence or secretory leader is operably linked to DNA for a polypeptide if it is expressed as a preprotein that participates in the secretion of the polypeptide; a promoter or enhancer is operably linked to a coding sequence if it affects the transcription of the sequence; or a ribosome binding site is operably linked to a coding sequence if it is positioned so as to facilitate translation. Generally, "operably linked" means that the DNA sequences being linked are contiguous, and, in the case of a secretory leader, contiguous and in reading phase. However, enhancers do not have to be contiguous. Linking is accomplished by ligation at convenient restriction sites. If such sites do not exist, the synthetic oligonucleotide adaptors or linkers are used in accordance with conventional practice.

The term "antibody" is used in the broadest sense and specifically covers single anti-PRO300 polypeptide monoclonal antibodies (including agonist, antagonist, and neutralizing antibodies) and anti-PRO300 antibody compositions with polyepitopic specificity. The term "monoclonal antibody" as used herein refers to an antibody obtained from a population of substantially homogeneous antibodies, *i.e.*, the individual antibodies comprising the population are identical except for possible naturally-occurring mutations that may be present in minor amounts.

"Active" or "activity" for the purposes herein refers to form(s) of PRO300 which retain the biologic and/or immunologic activities of native or naturally-occurring PRO300 polypeptide.

The terms "treating", "treatment" and "therapy" as used herein refer to curative therapy, prophylactic therapy, and preventative therapy.

The term "mammal" as used herein refers to any mammal classified as a mammal, including humans, cows, horses, dogs and cats. In a preferred embodiment of the invention, the mammal is a human.

II. Compositions and Methods of the Invention

A. Full-length PRO300 Polypeptide

The present invention provides newly identified and isolated nucleotide sequences encoding polypeptides referred to in the present application as PRO300. In particular, Applicants have identified and isolated cDNA encoding a PRO300 polypeptide, as disclosed in further detail in the Examples below. Using BLAST and FastA sequence alignment computer programs, Applicants found that various portions of the PRO300 polypeptide have significant homology with the human Diff 33 protein. Accordingly, it is presently believed that PRO300 polypeptide disclosed in the present application is a newly identified member of the Diff 33 family.

B. Modifications of PRO300

Covalent modifications of PRO300 polypeptides are included within the scope of this invention. One type of covalent modification includes reacting targeted amino acid residues of a PRO300 polypeptide with an organic derivatizing agent that is capable of reacting with selected side chains or the N-or C-terminal residues of a PRO300 polypeptide. Derivatization with bifunctional agents is useful, for instance, for crosslinking PRO300 to a water-insoluble support matrix or surface for use in the method for purifying anti-PRO300 antibodies, and vice-versa. Commonly used crosslinking agents include, *e.g.*, 1,1-bis(diazo-acetyl)-2-phenylethane, glutaraldehyde, N-hydroxysuccinimide esters, for example, esters with 4-azidosalicylic acid, homobifunctional imidoesters, including disuccinimidyl esters such as 3,3'-dithiobis(succinimidylpropionate), bifunctional maleimides such as bis-N-maleimido-1,8-octane and agents such as methyl-3-[(p-azidophenyl)dithio]propioimidate.

Other modifications include deamidation of glutaminyl and asparaginyl residues to the corresponding glutamyl and aspartyl residues, respectively, hydroxylation of proline and lysine, phosphorylation of hydroxyl groups of seryl or threonyl residues, methylation of the α-amino groups of lysine, arginine, and histidine side chains [T.E. Creighton, Proteins: Structure and Molecular Properties, W.H. Freeman & Co., San Francisco, pp. 79-86 (1983)], acetylation of the N-terminal amine, and amidation of any C-terminal carboxyl group.

Another type of covalent modification of the PRO300 polypeptide included within the scope of this invention comprises altering the native glycosylation pattern of the polypeptide. "Altering the native glycosylation pattern" is intended for purposes herein to mean deleting one

or more carbohydrate moieties found in native sequence PRO300 polypeptide, and/or adding one or more glycosylation sites that are not present in the native sequence PRO300 polypeptide.

Addition of glycosylation sites to PRO300 polypeptides may be accomplished by altering the amino acid sequence thereof. The alteration may be made, for example, by the addition of, or substitution by, one or more serine or threonine residues to the native sequence PRO300 polypeptide (for O-linked glycosylation sites). The PRO300 amino acid sequence may optionally be altered through changes at the DNA level, particularly by mutating the DNA encoding the PRO300 polypeptide at preselected bases such that codons are generated that will translate into the desired amino acids.

Another means of increasing the number of carbohydrate moieties on the PRO300 polypeptide is by chemical or enzymatic coupling of glycosides to the polypeptide. Such methods are described in the art, e.g., in WO 87/05330 published 11 September 1987, and in Aplin and Wriston, CRC Crit. Rev. Biochem., pp. 259-306 (1981).

Removal of carbohydrate moieties present on the PRO300 polypeptide may be accomplished chemically or enzymatically or by mutational substitution of codons encoding for amino acid residues that serve as targets for glycosylation. Chemical deglycosylation techniques are known in the art and described, for instance, by Hakimuddin, et al., Arch. Biochem. Biophys., 259:52 (1987) and by Edge et al., Anal. Biochem., 118:131 (1981). Enzymatic cleavage of carbohydrate moieties on polypeptides can be achieved by the use of a variety of endo-and exo-glycosidases as described by Thotakura et al., Meth. Enzymol., 138:350 (1987).

Another type of covalent modification of PRO300 comprises linking the PRO300 polypeptide to one of a variety of nonproteinaceous polymers, *e.g.*, polyethylene glycol, polypropylene glycol, or polyoxyalkylenes, in the manner set forth in U.S. Patent Nos. 4,640,835; 4,496,689; 4,301,144; 4,670,417; 4,791,192 or 4,179,337.

PRO300 polypeptides of the present invention may also be modified in a way to form chimeric molecules comprising a PRO300 polypeptide fused to another, heterologous polypeptide or amino acid sequence. In one embodiment, such a chimeric molecule comprises a fusion of a PRO300 polypeptide with a tag polypeptide which provides an epitope to which an anti-tag antibody can selectively bind. The epitope tag is generally placed at the amino-or carboxyl-terminus of the PRO300 polypeptide. The presence of such epitope-tagged forms of

a PRO300 polypeptide can be detected using an antibody against the tag polypeptide. Also, provision of the epitope tag enables the PRO300 polypeptide to be readily purified by affinity purification using an anti-tag antibody or another type of affinity matrix that binds to the epitope tag. In an alternative embodiment, the chimeric molecule may comprise a fusion of a PRO300 polypeptide with an immunoglobulin or a particular region of an immunoglobulin. For a bivalent form of the chimeric molecule, such a fusion could be to the Fc region of an IgG molecule.

Various tag polypeptides and their respective antibodies are well known in the art. Examples include poly-histidine (poly-his) or poly-histidine-glycine (poly-his-gly) tags; the flu HA tag polypeptide and its antibody 12CA5 [Field et al., Mol. Cell. Biol., 8:2159-2165 (1988)]; the c-myc tag and the 8F9, 3C7, 6E10, G4, B7 and 9E10 antibodies thereto [Evan et al., Molecular and Cellular Biology, 5:3610-3616 (1985)]; and the Herpes Simplex virus glycoprotein D (gD) tag and its antibody [Paborsky et al., Protein Engineering, 3(6):547-553 (1990)]. Other tag polypeptides include the Flag-peptide [Hopp et al., BioTechnology, 6:1204-1210 (1988)]; the KT3 epitope peptide [Martin et al., Science, 255:192-194 (1992)]; an α-tubulin epitope peptide [Skinner et al., J. Biol. Chem., 266:15163-15166 (1991)]; and the T7 gene 10 protein peptide tag [Lutz-Freyermuth et al., Proc. Natl. Acad. Sci. USA, 87:6393-6397 (1990)].

C. Preparation of PRO300

The description below relates primarily to production of PRO300 by culturing cells transformed or transfected with a vector containing at least nucleotides 45-1415 (SEQ ID NO:2) of the UNQ263 (DNA40625) nucleic acid (SEQ ID NO:1). It is, of course, contemplated that alternative methods, which are well known in the art, may be employed to prepare PRO300 polypeptides. For instance, the PRO300 sequence, or portions thereof, may be produced by direct peptide synthesis using solid-phase techniques [see, e.g., Stewart et al., Solid-Phase Peptide Synthesis, W.H. Freeman Co., San Francisco, CA (1969); Merrifield, J. Am. Chem. Soc., 85:2149-2154 (1963)]. *In vitro* protein synthesis may be performed using manual techniques or by automation. Automated synthesis may be accomplished, for instance, using an Applied Biosystems Peptide Synthesizer (Foster City, CA) using manufacturer's instructions. Various portions of PRO300 polypeptides may be chemically synthesized separately and combined using chemical or enzymatic methods to produce a full-length

PRO300 polypeptide.

1. Isolation of DNA Encoding PRO300

DNA encoding a PRO300 polypeptide may be obtained from a cDNA library prepared from tissue believed to possess the PRO300 mRNA and to express it at a detectable level. Accordingly, human PRO300-encoding DNA can be conveniently obtained from a cDNA library prepared from human tissue, such as described in the Examples. The PRO300-encoding gene may also be obtained from a genomic library or by oligonucleotide synthesis.

Libraries can be screened with probes (such as antibodies to a PRO300 polypeptide or oligonucleotides of at least about 20-80 bases) designed to identify the gene of interest or the protein encoded by it. Screening the cDNA or genomic library with the selected probe may be conducted using standard procedures, such as described in Sambrook et al., Molecular Cloning: A Laboratory Manual (New York: Cold Spring Harbor Laboratory Press, 1989). An alternative means to isolate the gene encoding PRO300 is to use PCR methodology [Sambrook et al., supra; Dieffenbach et al., PCR Primer:A Laboratory Manual (Cold Spring Harbor Laboratory Press, 1995)].

The Examples below describe techniques for screening a cDNA library. The oligonucleotide sequences selected as probes should be of sufficient length and sufficiently unambiguous that false positives are minimized. The oligonucleotide is preferably labeled such that it can be detected upon hybridization to DNA in the library being screened. Methods of labeling are well known in the art, and include the use of radiolabels like ^{32}P-labeled ATP, biotinylation or enzyme labeling. Hybridization conditions, including moderate stringency and high stringency, are provided in Sambrook et al., supra.

Sequences identified in such library screening methods can be compared and aligned to other known sequences deposited and available in public databases such as GenBank or other private sequence databases. Sequence identity (at either the amino acid or nucleotide level) within defined regions of the molecule or across the full-length sequence can be determined through sequence alignment using computer software programs such as ALIGN, DNAstar, and INHERIT which employ various algorithms to measure homology.

Nucleic acid having protein coding sequence may be obtained by screening selected cDNA or genomic libraries using the deduced amino acid sequence disclosed herein for the

first time, and, if necessary, using conventional primer extension procedures as described in Sambrook et al., supra, to detect precursors and processing intermediates of mRNA that may not have been reverse-transcribed into cDNA.

2. Selection and Transformation of Host Cells

Host cells are transfected or transformed with expression or cloning vectors described herein for PRO300 polypeptide production and cultured in conventional nutrient media modified as appropriate for inducing promoters, selecting transformants, or amplifying the genes encoding the desired sequences. The culture conditions, such as media, temperature, pH and the like, can be selected by the skilled artisan without undue experimentation. In general, principles, protocols, and practical techniques for maximizing the productivity of cell cultures can be found in Mammalian Cell Biotechnology: a Practical Approach, M. Butler, ed. (IRL Press, 1991) and Sambrook et al., supra.

Methods of transfection are known to the ordinarily skilled artisan, for example, $CaPO_4$ and electroporation. Depending on the host cell used, transformation is performed using standard techniques appropriate to such cells. The calcium treatment employing calcium chloride, as described in Sambrook et al., supra, or electroporation is generally used for prokaryotes or other cells that contain substantial cell-wall barriers. Infection with *Agrobacterium tumefaciens* is used for transformation of certain plant cells, as described by Shaw et al., Gene, 23:315 (1983) and WO 89/05859 published 29 June 1989. For mammalian cells without such cell walls, the calcium phosphate precipitation method of Graham and van der Eb, Virology, 52:456-457 (1978) can be employed. General aspects of mammalian cell host system transformations have been described in U.S. Patent No. 4,399,216. Transformations into yeast are typically carried out according to the method of Van Solingen et al., J. Bact., 130:946 (1977) and Hsiao et al., Proc. Natl. Acad. Sci. (USA), 76:3829 (1979). However, other methods for introducing DNA into cells, such as by nuclear microinjection, electroporation, bacterial protoplast fusion with intact cells, or polycations, *e.g.*, polybrene, polyornithine, may also be used. For various techniques for transforming mammalian cells, see Keown et al., Methods in Enzymology, 185:527-537 (1990) and Mansour et al., Nature, 336:348-352 (1988).

Suitable host cells for cloning or expressing the DNA in the vectors herein include prokaryote, yeast, or higher eukaryote cells. Suitable prokaryotes include but are not limited

to eubacteria, such as Gram-negative or Gram-positive organisms, for example, Enterobacteriaceae such as *E. coli*. Various *E. coli* strains are publicly available, such as *E. coli* K12 strain MM294 (ATCC 31,446); *E. coli* X1776 (ATCC 31,537); *E. coli* strain W3110 (ATCC 27,325) and K5 772 (ATCC 53,635).

In addition to prokaryotes, eukaryotic microbes such as filamentous fungi or yeast are suitable cloning or expression hosts for PRO300-encoding vectors. *Saccharomyces cerevisiae* is a commonly used lower eukaryotic host microorganism.

Suitable host cells for the expression of glycosylated PRO300 are derived from multicellular organisms. Examples of invertebrate cells include insect cells such as Drosophila S2 and Spodoptera Sf9, as well as plant cells. Examples of useful mammalian host cell lines include Chinese hamster ovary (CHO) and COS cells. More specific examples include monkey kidney CV1 line transformed by SV40 (COS-7, ATCC CRL 1651); human embryonic kidney line (293 or 293 cells subcloned for growth in suspension culture, Graham et al., J. Gen Virol., 36:59 (1977)); Chinese hamster ovary cells/-DHFR (CHO, Urlaub and Chasin, Proc. Natl. Acad. Sci. USA, 77:4216 (1980)); mouse sertoli cells (TM4, Mather, Biol. Reprod., 23:243-251 (1980)); human lung cells (W138, ATCC CCL 75); human liver cells (Hep G2, HB 8065); and mouse mammary tumor (MMT 060562, ATCC CCL51). The selection of the appropriate host cell is deemed to be within the skill in the art.

3. Selection and Use of a Replicable Vector

The nucleic acid (*e.g.*, cDNA or genomic DNA) encoding the desired PRO300 polypeptide may be inserted into a replicable vector for cloning (amplification of the DNA) or for expression. Various vectors are publicly available. The vector may, for example, be in the form of a plasmid, cosmid, viral particle, or phage. The appropriate nucleic acid sequence may be inserted into the vector by a variety of procedures. In general, DNA is inserted into an appropriate restriction endonuclease site(s) using techniques known in the art. Vector components generally include, but are not limited to, one or more of a signal sequence, an origin of replication, one or more marker genes, an enhancer element, a promoter, and a transcription termination sequence. Construction of suitable vectors containing one or more of these components employs standard ligation techniques which are known to the skilled artisan.

The desired PRO300 polypeptide may be produced recombinantly not only directly, but also as a fusion polypeptide with a heterologous polypeptide, which may be a signal sequence or other polypeptide having a specific cleavage site at the N-terminus of the mature protein or polypeptide. In general, the signal sequence may be a component of the vector, or it may be a part of the PRO300-encoding DNA that is inserted into the vector. The signal sequence may be a prokaryotic signal sequence selected, for example, from the group of the alkaline phosphatase, penicillinase, lpp, or heat-stable enterotoxin II leaders. For yeast secretion the signal sequence may be, *e.g.*, the yeast invertase leader, alpha factor leader (including *Saccharomyces* and *Kluyveromyces* α-factor leaders, the latter described in U.S. Patent No. 5,010,182), or acid phosphatase leader, the *C. albicans* glucoamylase leader (EP 362,179 published 4 April 1990), or the signal described in WO 90/13646 published 15 November 1990. In mammalian cell expression, mammalian signal sequences may be used to direct secretion of the protein, such as signal sequences from secreted polypeptides of the same or related species, as well as viral secretory leaders.

Both expression and cloning vectors contain a nucleic acid sequence that enables the vector to replicate in one or more selected host cells. Such sequences are well known for a variety of bacteria, yeast, and viruses. The origin of replication from the plasmid pBR322 is suitable for most Gram-negative bacteria, the 2μ plasmid origin is suitable for yeast, and various viral origins (SV40, polyoma, adenovirus, VSV or BPV) are useful for cloning vectors in mammalian cells.

Expression and cloning vectors will typically contain a selection gene, also termed a selectable marker. Typical selection genes encode proteins that (a) confer resistance to antibiotics or other toxins, *e.g.*, ampicillin, neomycin, methotrexate, or tetracycline, (b) complement auxotrophic deficiencies, or (c) supply critical nutrients not available from complex media, *e.g.*, the gene encoding D-alanine racemase for *Bacilli*.

An example of suitable selectable markers for mammalian cells are those that enable the identification of cells competent to take up the PRO300-encoding nucleic acid, such as DHFR or thymidine kinase. An appropriate host cell when wild-type DHFR is employed is the CHO cell line deficient in DHFR activity, prepared and propagated as described by Urlaub et al., Proc. Natl. Acad. Sci. USA, 77:4216 (1980). A suitable selection gene for use in yeast is the *trp*1 gene present in the yeast plasmid YRp7 [Stinchcomb et al., Nature, 282:39 (1979); Kingsman et al., Gene, 7:141 (1979); Tschemper et al., Gene, 10:157 (1980)]. The *trp*1 gene

provides a selection marker for a mutant strain of yeast lacking the ability to grow in tryptophan, for example, ATCC No. 44076 or PEP4-1 [Jones, Genetics, 85:12 (1977)].

Expression and cloning vectors usually contain a promoter operably linked to the PRO300-encoding nucleic acid sequence to direct mRNA synthesis. Promoters recognized by a variety of potential host cells are well known. Promoters suitable for use with prokaryotic hosts include the β-lactamase and lactose promoter systems [Chang et al., Nature, 275:615 (1978); Goeddel et al., Nature, 281:544 (1979)], alkaline phosphatase, a tryptophan (trp) promoter system [Goeddel, Nucleic Acids Res., 8:4057 (1980); EP 36,776], and hybrid promoters such as the tac promoter [deBoer et al., Proc. Natl. Acad. Sci. USA, 80:21-25 (1983)]. Promoters for use in bacterial systems also will contain a Shine-Dalgarno (S.D.) sequence operably linked to the DNA encoding the PRO300 polypeptide.

Examples of suitable promoting sequences for use with yeast hosts include the promoters for 3-phosphoglycerate kinase [Hitzeman et al., J. Biol. Chem., 255:2073 (1980)] or other glycolytic enzymes [Hess et al., J. Adv. Enzyme Reg., 7:149 (1968); Holland, Biochemistry, 17:4900 (1978)], such as enolase, glyceraldehyde-3-phosphate dehydrogenase, hexokinase, pyruvate decarboxylase, phosphofructokinase, glucose-6-phosphate isomerase, 3-phosphoglycerate mutase, pyruvate kinase, triosephosphate isomerase, phosphoglucose isomerase, and glucokinase.

Other yeast promoters, which are inducible promoters having the additional advantage of transcription controlled by growth conditions, are the promoter regions for alcohol dehydrogenase 2, isocytochrome C, acid phosphatase, degradative enzymes associated with nitrogen metabolism, metallothionein, glyceraldehyde-3-phosphate dehydrogenase, and enzymes responsible for maltose and galactose utilization. Suitable vectors and promoters for use in yeast expression are further described in EP 73,657.

PRO300 transcription from vectors in mammalian host cells is controlled, for example, by promoters obtained from the genomes of viruses such as polyoma virus, fowlpox virus (UK 2,211,504 published 5 July 1989), adenovirus (such as Adenovirus 2), bovine papilloma virus, avian sarcoma virus, cytomegalovirus, a retrovirus, hepatitis-B virus and Simian Virus 40 (SV40), from heterologous mammalian promoters, *e.g.*, the actin promoter or an immunoglobulin promoter, and from heat-shock promoters, provided such promoters are compatible with the host cell systems.

Transcription of a DNA encoding a PRO300 polypeptide by higher eukaryotes may be increased by inserting an enhancer sequence into the vector. Enhancers are cis-acting elements of DNA, usually about from 10 to 300 bp, that act on a promoter to increase its transcription. Many enhancer sequences are now known from mammalian genes (globin, elastase, albumin, α-fetoprotein, and insulin). Typically, however, one will use an enhancer from a eukaryotic cell virus. Examples include the SV40 enhancer on the late side of the replication origin (bp 100-270), the cytomegalovirus early promoter enhancer, the polyoma enhancer on the late side of the replication origin, and adenovirus enhancers. The enhancer may be spliced into the vector at a position 5' or 3' to the PRO300 coding sequence, but is preferably located at a site 5' from the promoter.

Expression vectors used in eukaryotic host cells (yeast, fungi, insect, plant, animal, human, or nucleated cells from other multicellular organisms) will also contain sequences necessary for the termination of transcription and for stabilizing the mRNA. Such sequences are commonly available from the 5' and, occasionally 3', untranslated regions of eukaryotic or viral DNAs or cDNAs. These regions contain nucleotide segments transcribed as polyadenylated fragments in the untranslated portion of the mRNA encoding PRO300.

Still other methods, vectors, and host cells suitable for adaptation to the synthesis of PRO300 polypeptides in recombinant vertebrate cell culture are described in Gething et al., Nature, 293:620-625 (1981); Mantei et al., Nature, 281:40-46 (1979); EP 117,060; and EP 117,058.

4. Detecting Gene Amplification/Expression

Gene amplification and/or expression may be measured in a sample directly, for example, by conventional Southern blotting, Northern blotting to quantitate the transcription of mRNA [Thomas, Proc. Natl. Acad. Sci. USA, 77:5201-5205 (1980)], dot blotting (DNA analysis), or *in situ* hybridization, using an appropriately labeled probe, based on the sequences provided herein. Alternatively, antibodies may be employed that can recognize specific duplexes, including DNA duplexes, RNA duplexes, and DNA-RNA hybrid duplexes or DNA-protein duplexes. The antibodies in turn may be labeled and the assay may be carried out where the duplex is bound to a surface, so that upon the formation of duplex on the surface, the presence of antibody bound to the duplex can be detected.

Gene expression, alternatively, may be measured by immunological methods, such as immunohistochemical staining of cells or tissue sections and assay of cell culture or body fluids, to quantitate directly the expression of gene product. Antibodies useful for immunohistochemical staining and/or assay of sample fluids may be either monoclonal or polyclonal, and may be prepared in any mammal. Conveniently, the antibodies may be prepared against a native sequence PRO300 polypeptide or against a synthetic peptide based on the DNA sequences provided herein or against exogenous sequence fused to PRO300-encoding DNA and encoding a specific antibody epitope.

5. Purification of Polypeptide

Forms of PRO300 may be recovered from culture medium or from host cell lysates. If membrane-bound, it can be released from the membrane using a suitable detergent solution (*e.g.* Triton-X 100) or by enzymatic cleavage. Cells employed in expression of PRO300 polypeptides can be disrupted by various physical or chemical means, such as freeze-thaw cycling, sonication, mechanical disruption, or cell lysing agents.

It may be desired to purify PRO300 from recombinant cell proteins or polypeptides. The following procedures are exemplary of suitable purification procedures: by fractionation on an ion-exchange column; ethanol precipitation; reverse phase HPLC; chromatography on silica or on a cation-exchange resin such as DEAE; chromatofocusing; SDS-PAGE; ammonium sulfate precipitation; gel filtration using, for example, Sephadex G-75; protein A Sepharose columns to remove contaminants such as IgG; and metal chelating columns to bind epitope-tagged forms of the PRO300 polypeptide. Various methods of protein purification may be employed and such methods are known in the art and described for example in Deutscher, Methods in Enzymology, 182 (1990); Scopes, Protein Purification:Principles and Practice, Springer-Verlag, New York (1982). The purification step(s) selected will depend, for example, on the nature of the production process used and the particular PRO300 polypeptide produced.

D. Uses for PRO300

Nucleotide sequences (or their complement) encoding PRO300 polypeptides have various applications in the art of molecular biology, including uses as hybridization probes, in chromosome and gene mapping and in the generation of anti-sense RNA and DNA.

PRO300-encoding nucleic acid will also be useful for the preparation of PRO300 polypeptides by the recombinant techniques described herein.

The full-length nucleotide sequence SEQ ID NO:1 or the full-length native sequence PRO240 (SEQ ID NO:2) nucleotide sequence, or portions thereof, may be used as hybridization probes for a cDNA library to isolate the full-length PRO300 gene or to isolate still other genes (for instance, those encoding naturally-occurring variants of PRO300 or PRO300 from other species) which have a desired sequence identity to the PRO300 sequence disclosed in Figure 1 (SEQ ID NO:1). Optionally, the length of the probes will be about 20 to about 50 bases. The hybridization probes may be derived from the UNQ263 (DNA40625-1189) nucleotide sequence of SEQ ID NO:1 as shown in Figure 1 or from genomic sequences including promoters, enhancer elements and introns of native sequence PRO300-encoding DNA. By way of example, a screening method will comprise isolating the coding region of the PRO300 gene using the known DNA sequence to synthesize a selected probe of about 40 bases. Hybridization probes may be labeled by a variety of labels, including radionucleotides such as ^{32}P or ^{35}S, or enzymatic labels such as alkaline phosphatase coupled to the probe via avidin/biotin coupling systems. Labeled probes having a sequence complementary to that of the PRO300 gene of the present invention can be used to screen libraries of human cDNA, genomic DNA or mRNA to determine which members of such libraries the probe hybridizes to. Hybridization techniques are described in further detail in the Examples below.

The probes may also be employed in PCR techniques to generate a pool of sequences for identification of closely related PRO300 sequences.

Nucleotide sequences encoding a PRO300 polypeptide can also be used to construct hybridization probes for mapping the gene which encodes that PRO300 polypeptide and for the genetic analysis of individuals with genetic disorders. The nucleotide sequences provided herein may be mapped to a chromosome and specific regions of a chromosome using known techniques, such as *in situ* hybridization, linkage analysis against known chromosomal markers, and hybridization screening with libraries.

When the coding sequences for PRO300 encode a protein which binds to another protein (example, where the PRO300 polypeptide functions as a receptor), the PRO300 polypeptide can be used in assays to identify the other proteins or molecules involved in the binding interaction. By such methods, inhibitors of the receptor/ligand binding interaction can be identified. Proteins involved in such binding interactions can also be used to screen for

peptide or small molecule inhibitors or agonists of the binding interaction. Also, the receptor PRO300 polypeptide can be used to isolate correlative ligand(s). Screening assays can be designed to find lead compounds that mimic the biological activity of a native PRO300 or a receptor for PRO300. Such screening assays will include assays amenable to high-throughput screening of chemical libraries, making them particularly suitable for identifying small molecule drug candidates. Small molecules contemplated include synthetic organic or inorganic compounds. The assays can be performed in a variety of formats, including protein-protein binding assays, biochemical screening assays, immunoassays and cell based assays, which are well characterized in the art.

Nucleic acids which encode PRO300 polypeptide or any of its modified forms can also be used to generate either transgenic animals or "knock out" animals which, in turn, are useful in the development and screening of therapeutically useful reagents. A transgenic animal (e.g., a mouse or rat) is an animal having cells that contain a transgene, which transgene was introduced into the animal or an ancestor of the animal at a prenatal, e.g., an embryonic stage. A transgene is a DNA which is integrated into the genome of a cell from which a transgenic animal develops. In one embodiment, cDNA encoding PRO300 polypeptide can be used to clone genomic DNA encoding PRO300 in accordance with established techniques and the genomic sequences used to generate transgenic animals that contain cells which express DNA encoding PRO300. Methods for generating transgenic animals, particularly animals such as mice or rats, have become conventional in the art and are described, for example, in U.S. Patent Nos. 4,736,866 and 4,870,009. Typically, particular cells would be targeted for PRO300 transgene incorporation with tissue-specific enhancers. Transgenic animals that include a copy of a transgene encoding PRO300 introduced into the germ line of the animal at an embryonic stage can be used to examine the effect of increased expression of DNA encoding PRO300. Such animals can be used as tester animals for reagents thought to confer protection from, for example, pathological conditions associated with its overexpression. In accordance with this facet of the invention, an animal is treated with the reagent and a reduced incidence of the pathological condition, compared to untreated animals bearing the transgene, would indicate a potential therapeutic intervention for the pathological condition.

Alternatively, non-human homologues of PRO300 can be used to construct a PRO300 "knock out" animal which has a defective or altered gene encoding PRO300 as a result of homologous recombination between the endogenous gene encoding PRO300 and altered

genomic DNA encoding PRO300 introduced into an embryonic cell of the animal. For example, cDNA encoding PRO300 can be used to clone genomic DNA encoding PRO300 in accordance with established techniques. A portion of the genomic DNA encoding PRO300 can be deleted or replaced with another gene, such as a gene encoding a selectable marker which can be used to monitor integration. Typically, several kilobases of unaltered flanking DNA (both at the 5' and 3' ends) are included in the vector [see e.g., Thomas and Capecchi, Cell, 51:503 (1987) for a description of homologous recombination vectors]. The vector is introduced into an embryonic stem cell line (e.g., by electroporation) and cells in which the introduced DNA has homologously recombined with the endogenous DNA are selected [see e.g., Li et al., Cell, 69:915 (1992)]. The selected cells are then injected into a blastocyst of an animal (e.g., a mouse or rat) to form aggregation chimeras [see e.g., Bradley, in *Teratocarcinomas and Embryonic Stem Cells: A Practical Approach*, E. J. Robertson, ed. (IRL, Oxford, 1987), pp. 113-152]. A chimeric embryo can then be implanted into a suitable pseudopregnant female foster animal and the embryo brought to term to create a "knock out" animal. Progeny harboring the homologously recombined DNA in their germ cells can be identified by standard techniques and used to breed animals in which all cells of the animal contain the homologously recombined DNA. Knockout animals can be characterized for instance, for their ability to defend against certain pathological conditions and for their development of pathological conditions due to absence of the PRO300 polypeptide.

PRO300 polypeptides and portions thereof which have homology to Diff 33 may also be useful for *in vivo* therapeutic purposes, as well as for various other applications. The identification of novel Diff 33-like molecules may have relevance to a number of human disorders such as the physiology of cancer. Thus, the identification of new Diff 33-like molecules is of special importance in that such proteins may serve as potential therapeutics for a variety of different human disorders. Such polypeptides may also play important roles in biotechnological and medical research as well as various industrial applications. As a result, there is particular scientific and medical interest in new molecules, such as PRO300.

E. Anti-PRO300 Antibodies

The present invention further provides anti-PRO300 polypeptide antibodies. Exemplary antibodies include polyclonal, monoclonal, humanized, bispecific, and heteroconjugate antibodies.

1. Polyclonal Antibodies

The anti-PRO300 antibodies of the present invention may comprise polyclonal antibodies. Methods of preparing polyclonal antibodies are known to the skilled artisan. Polyclonal antibodies can be raised in a mammal, for example, by one or more injections of an immunizing agent and, if desired, an adjuvant. Typically, the immunizing agent and/or adjuvant will be injected in the mammal by multiple subcutaneous or intraperitoneal injections. The immunizing agent may include the PRO300 polypeptide or a fusion protein thereof. It may be useful to conjugate the immunizing agent to a protein known to be immunogenic in the mammal being immunized. Examples of such immunogenic proteins include but are not limited to keyhole limpet hemocyanin, serum albumin, bovine thyroglobulin, and soybean trypsin inhibitor. Examples of adjuvants which may be employed include Freund's complete adjuvant and MPL-TDM adjuvant (monophosphoryl Lipid A, synthetic trehalose dicorynomycolate). The immunization protocol may be selected by one skilled in the art without undue experimentation.

2. Monoclonal Antibodies

The anti-PRO300 antibodies may, alternatively, be monoclonal antibodies. Monoclonal antibodies may be prepared using hybridoma methods, such as those described by Kohler and Milstein, Nature, 256:495 (1975). In a hybridoma method, a mouse, hamster, or other appropriate host animal, is typically immunized with an immunizing agent to elicit lymphocytes that produce or are capable of producing antibodies that will specifically bind to the immunizing agent. Alternatively, the lymphocytes may be immunized *in vitro*.

The immunizing agent will typically include the PRO300 polypeptide or a fusion protein thereof. Generally, either peripheral blood lymphocytes ("PBLs") are used if cells of human origin are desired, or spleen cells or lymph node cells are used if non-human mammalian sources are desired. The lymphocytes are then fused with an immortalized cell line using a suitable fusing agent, such as polyethylene glycol, to form a hybridoma cell [Goding, Monoclonal Antibodies: Principles and Practice, Academic Press, (1986) pp. 59-103]. Immortalized cell lines are usually transformed mammalian cells, particularly myeloma cells of rodent, bovine and human origin. Usually, rat or mouse myeloma cell lines are employed. The hybridoma cells may be cultured in a suitable culture medium that preferably contains one or more substances that inhibit the growth or survival of the unfused,

immortalized cells. For example, if the parental cells lack the enzyme hypoxanthine guanine phosphoribosyl transferase (HGPRT or HPRT), the culture medium for the hybridomas typically will include hypoxanthine, aminopterin, and thymidine ("HAT medium"), which substances prevent the growth of HGPRT-deficient cells.

Preferred immortalized cell lines are those that fuse efficiently, support stable high level expression of antibody by the selected antibody-producing cells, and are sensitive to a medium such as HAT medium. More preferred immortalized cell lines are murine myeloma lines, which can be obtained, for instance, from the Salk Institute Cell Distribution Center, San Diego, California and the American Type Culture Collection, Rockville, Maryland. Human myeloma and mouse-human heteromyeloma cell lines also have been described for the production of human monoclonal antibodies [Kozbor, J. Immunol., 133:3001 (1984); Brodeur et al., Monoclonal Antibody Production Techniques and Applications, Marcel Dekker, Inc., New York, (1987) pp. 51-63].

The culture medium in which the hybridoma cells are cultured can then be assayed for the presence of monoclonal antibodies directed against a PRO300 polypeptide. Preferably, the binding specificity of monoclonal antibodies produced by the hybridoma cells is determined by immunoprecipitation or by an *in vitro* binding assay, such as radioimmunoassay (RIA) or enzyme-linked immunoabsorbent assay (ELISA). Such techniques and assays are known in the art. The binding affinity of the monoclonal antibody can, for example, be determined by the Scatchard analysis of Munson and Pollard, Anal. Biochem., 107:220 (1980).

After the desired hybridoma cells are identified, the clones may be subcloned by limiting dilution procedures and grown by standard methods [Goding, supra]. Suitable culture media for this purpose include, for example, Dulbecco's Modified Eagle's Medium and RPMI-1640 medium. Alternatively, the hybridoma cells may be grown *in vivo* as ascites in a mammal.

The monoclonal antibodies secreted by the subclones may be isolated or purified from the culture medium or ascites fluid by conventional immunoglobulin purification procedures such as, for example, protein A-Sepharose, hydroxylapatite chromatography, gel electrophoresis, dialysis, or affinity chromatography.

The monoclonal antibodies may also be made by recombinant DNA methods, such as those described in U.S. Patent No. 4,816,567. DNA encoding the monoclonal antibodies of the invention can be readily isolated and sequenced using conventional procedures (e.g., by

using oligonucleotide probes that are capable of binding specifically to genes encoding the heavy and light chains of murine antibodies). The hybridoma cells of the invention serve as a preferred source of such DNA. Once isolated, the DNA may be placed into expression vectors, which are then transfected into host cells such as simian COS cells, Chinese hamster ovary (CHO) cells, or myeloma cells that do not otherwise produce immunoglobulin protein, to obtain the synthesis of monoclonal antibodies in the recombinant host cells. The DNA also may be modified, for example, by substituting the coding sequence for human heavy and light chain constant domains in place of the homologous murine sequences [U.S. Patent No. 4,816,567; Morrison et al., supra] or by covalently joining to the immunoglobulin coding sequence all or part of the coding sequence for a non-immunoglobulin polypeptide. Such a non-immunoglobulin polypeptide can be substituted for the constant domains of an antibody of the invention, or can be substituted for the variable domains of one antigen-combining site of an antibody of the invention to create a chimeric bivalent antibody.

The antibodies may be monovalent antibodies. Methods for preparing monovalent antibodies are well known in the art. For example, one method involves recombinant expression of immunoglobulin light chain and modified heavy chain. The heavy chain is truncated generally at any point in the Fc region so as to prevent heavy chain crosslinking. Alternatively, the relevant cysteine residues are substituted with another amino acid residue or are deleted so as to prevent crosslinking.

In vitro methods are also suitable for preparing monovalent antibodies. Digestion of antibodies to produce fragments thereof, particularly, Fab fragments, can be accomplished using routine techniques known in the art.

3. Humanized Antibodies

The anti-PRO300 antibodies of the invention may further comprise humanized antibodies or human antibodies. Humanized forms of non-human (e.g., murine) antibodies are chimeric immunoglobulins, immunoglobulin chains or fragments thereof (such as Fv, Fab, Fab', $F(ab')_2$ or other antigen-binding subsequences of antibodies) which contain minimal sequence derived from non-human immunoglobulin. Humanized antibodies include human immunoglobulins (recipient antibody) in which residues from a complementary determining region (CDR) of the recipient are replaced by residues from a CDR of a non-human species (donor antibody) such as mouse, rat or rabbit having the desired specificity, affinity and

capacity. In some instances, Fv framework residues of the human immunoglobulin are replaced by corresponding non-human residues. Humanized antibodies may also comprise residues which are found neither in the recipient antibody nor in the imported CDR or framework sequences. In general, the humanized antibody will comprise substantially all of at least one, and typically two, variable domains, in which all or substantially all of the CDR regions correspond to those of a non-human immunoglobulin and all or substantially all of the FR regions are those of a human immunoglobulin consensus sequence. The humanized antibody optimally also will comprise at least a portion of an immunoglobulin constant region (Fc), typically that of a human immunoglobulin [Jones et al., Nature, 321:522-525 (1986); Riechmann et al., Nature, 332:323-329 (1988); and Presta, Curr. Op. Struct. Biol., 2:593-596 (1992)].

Methods for humanizing non-human antibodies are well known in the art. Generally, a humanized antibody has one or more amino acid residues introduced into it from a source which is non-human. These non-human amino acid residues are often referred to as "import" residues, which are typically taken from an "import" variable domain. Humanization can be essentially performed following the method of Winter and co-workers [Jones et al., Nature, 321:522-525 (1986); Riechmann et al., Nature, 332:323-327 (1988); Verhoeyen et al., Science, 239:1534-1536 (1988)], by substituting rodent CDRs or CDR sequences for the corresponding sequences of a human antibody. Accordingly, such "humanized" antibodies are chimeric antibodies (U.S. Patent No. 4,816,567), wherein substantially less than an intact human variable domain has been substituted by the corresponding sequence from a non-human species. In practice, humanized antibodies are typically human antibodies in which some CDR residues and possibly some FR residues are substituted by residues from analogous sites in rodent antibodies.

Human antibodies can also be produced using various techniques known in the art, including phage display libraries [Hoogenboom and Winter, J. Mol. Biol., 227:381 (1991); Marks et al., J. Mol. Biol., 222:581 (1991)]. The techniques of Cole et al. and Boerner et al. are also available for the preparation of human monoclonal antibodies (Cole et al., Monoclonal Antibodies and Cancer Therapy, Alan R. Liss, p. 77 (1985) and Boerner et al., J. Immunol., 147(1):86-95 (1991)].

4. Bispecific Antibodies

Bispecific antibodies are monoclonal, preferably human or humanized, antibodies that have binding specificities for at least two different antigens. In the present case, one of the binding specificities is for a PRO300 polypeptide, the other one is for any other antigen, and preferably for a cell-surface protein or receptor or receptor subunit.

Methods for making bispecific antibodies are known in the art. Traditionally, the recombinant production of bispecific antibodies is based on the co-expression of two immunoglobulin heavy-chain/light-chain pairs, where the two heavy chains have different specificities [Milstein and Cuello, Nature, 305:537-539 (1983)]. Because of the random assortment of immunoglobulin heavy and light chains, these hybridomas (quadromas) produce a potential mixture of ten different antibody molecules, of which only one has the correct bispecific structure. The purification of the correct molecule is usually accomplished by affinity chromatography steps. Similar procedures are disclosed in WO 93/08829, published 13 May 1993, and in Traunecker et al., EMBO J., 10:3655-3659 (1991).

Antibody variable domains with the desired binding specificities (antibody-antigen combining sites) can be fused to immunoglobulin constant domain sequences. The fusion preferably is with an immunoglobulin heavy-chain constant domain, comprising at least part of the hinge, CH2, and CH3 regions. It is preferred to have the first heavy-chain constant region (CH1) containing the site necessary for light-chain binding present in at least one of the fusions. DNAs encoding the immunoglobulin heavy-chain fusions and, if desired, the immunoglobulin light chain, are inserted into separate expression vectors, and are co-transfected into a suitable host organism. For further details of generating bispecific antibodies see, for example, Suresh et al., Methods in Enzymology, 121:210 (1986).

5. Heteroconjugate Antibodies

Heteroconjugate antibodies are also within the scope of the present invention. Heteroconjugate antibodies are composed of two covalently joined antibodies. Such antibodies have, for example, been proposed to target immune system cells to unwanted cells [U.S. Patent No. 4,676,980], and for treatment of HIV infection [WO 91/00360; WO 92/200373; EP 03089]. It is contemplated that the antibodies may be prepared *in vitro* using known methods in synthetic protein chemistry, including those involving crosslinking agents. For example, immunotoxins may be constructed using a disulfide exchange reaction or by forming

a thioether bond. Examples of suitable reagents for this purpose include iminothiolate and methyl-4-mercaptobutyrimidate and those disclosed, for example, in U.S. Patent No. 4,676,980.

F. Uses for anti-PRO300 Antibodies

The anti-PRO300 antibodies of the present invention have various utilities. For example, anti-PRO300 antibodies may be used in diagnostic assays for PRO300 polypeptides, *e.g.*, detecting expression in specific cells, tissues, or serum. Various diagnostic assay techniques known in the art may be used, such as competitive binding assays, direct or indirect sandwich assays and immunoprecipitation assays conducted in either heterogeneous or homogeneous phases [Zola, Monoclonal Antibodies: A Manual of Techniques, CRC Press, Inc. (1987) pp. 147-158]. The antibodies used in the diagnostic assays can be labeled with a detectable moiety. The detectable moiety should be capable of producing, either directly or indirectly, a detectable signal. For example, the detectable moiety may be a radioisotope, such as ^{3}H, ^{14}C, ^{32}P, ^{35}S, or ^{125}I, a fluorescent or chemiluminescent compound, such as fluorescein isothiocyanate, rhodamine, or luciferin, or an enzyme, such as alkaline phosphatase, beta-galactosidase or horseradish peroxidase. Any method known in the art for conjugating the antibody to the detectable moiety may be employed, including those methods described by Hunter et al., Nature, 144:945 (1962); David et al., Biochemistry, 13:1014 (1974); Pain et al., J. Immunol. Meth., 40:219 (1981); and Nygren, J. Histochem. and Cytochem., 30:407 (1982).

Anti-PRO300 antibodies also are useful for the affinity purification of PRO300 polypeptides from recombinant cell culture or natural sources. In this process, the antibodies against a PRO300 polypeptide are immobilized on a suitable support, such a Sephadex resin or filter paper, using methods well known in the art. The immobilized antibody then is contacted with a sample containing the PRO300 polypeptide to be purified, and thereafter the support is washed with a suitable solvent that will remove substantially all the material in the sample except the PRO300 polypeptide, which is bound to the immobilized antibody. Finally, the support is washed with another suitable solvent that will release the PRO300 polypeptide from the antibody.

The following examples are offered for illustrative purposes only, and are not intended to limit the scope of the present invention in any way.

All patent and literature references cited in the present specification are hereby incorporated by reference in their entirety.

EXAMPLES

Commercially available reagents referred to in the examples were used according to manufacturer's instructions unless otherwise indicated. The source of those cells identified in the following examples, and throughout the specification, by ATCC accession numbers is the American Type Culture Collection, Rockville, Maryland.

EXAMPLE 1: Isolation of cDNA Clones Encoding Human PRO300

The extracellular domain (ECD) sequences (including the secretion signal, if any) of from about 950 known secreted proteins from the Swiss-Prot public protein database were used to search expressed sequence tag (EST) databases. The EST databases included public EST databases (e.g., GenBank) and a proprietary EST DNA database (LIFESEQ™, Incyte Pharmaceuticals, Palo Alto, CA). The search was performed using the computer program BLAST or BLAST2 (Altshul et al., Methods in Enzymology 266:460-480 (1996)) as a comparison of the ECD protein sequences to a 6 frame translation of the EST sequence. Those comparisons resulting in a BLAST score of 70 (or in some cases 90) or greater that did not encode known proteins were clustered and assembled into consensus DNA sequences with the program "phrap" (Phil Green, University of Washington, Seattle, Washington; http://bozeman.mbt.washington.edu/phrap.docs/phrap.html).

A consensus DNA sequence was assembled relative to other EST sequences using phrap. This consensus sequence is herein designated DNA35930 (see Figures 3A and 3B). The DNA35930 consensus DNA sequence was then extended using repeated cycles of BLAST and phrap to extend the consensus sequence as far as possible using the sources of EST sequences discussed above. The extended assembly sequence is indicated in numerous second alignment figures, as shown in Figures 4A-4AH, wherein the consensus sequence is therein designated <consen01>.

Based on the DNA35930 and <consen01> consensus sequences, oligonucleotides were synthesized: 1) to identify by PCR a cDNA library that contained the sequence of interest, and 2) for use as probes to isolate a clone of the full-length coding sequence for PRO300. Forward and reverse PCR primers generally range from 20 to 30 nucleotides and

are often designed to give a PCR product of about 100-1000 bp in length. The probe sequences are typically 40-55 bp in length. In some cases, additional oligonucleotides are synthesized when the consensus sequence is greater than about 1-1.5kbp. In order to screen several libraries for a full-length clone, DNA from the libraries was screened by PCR amplification, as ber Ausubel et al., Current Protocols in Molecular Biology, with the PCR primer pair. A positive library was then used to isolate clones encoding the gene of interest by the *in vivo* cloning procedure using the probe oligonucleotide and one of the primer pairs.

Forward and reverse PCR primers were synthesized:

forward PCR primer 1 (35930.f1) 5'-GCCGCCTCATCTTCACGTTCTTCC-3' (SEQ ID NO:4)

forward PCR primer 2 (35930.f2) 5'-TCATCCAGCTGGTGCTGCTC-3' (SEQ ID NO:5)

forward PCR primer 3 (35930.f3) 5'-CTTCTTCCACTTCTGCCTGG-3' (SEQ ID NO:6)

forward PCR primer 4 (35930.f4) 5'-CCTGGGCAAAAATGCAAC-3' (SEQ ID NO:7)

reverse PCR primer 1 (35930.r1) 5'-CAGGAATGTAGAAGGCACCCACGG-3' (SEQ ID NO:8)

reverse PCR primer 2 (35930.r2) 5'-TGGCACAGATCTTCACCCACACGG-3' (SEQ ID NO:9)

Additionally, a synthetic oligonucleotide hybridization probe was constructed from the consensus DNA35930 sequence which had the following nucleotide sequence

hybridization probe (35930.p1)
5'-TGTCCATCATTATGCTGAGCCCGGGCGTGGAGAGTCAGCTCTACAAGCTG-3' (SEQ ID NO:10)

In order to screen several libraries for a source of a full-length clone, DNA from the libraries was screened by PCR amplification with the PCR primer pair identified above. A positive library was then used to isolate clones encoding the PRO300 gene using the probe oligonucleotide and one of the PCR primers.

RNA for construction of the cDNA libraries was isolated from human fetal kidney tissue. The cDNA libraries used to isolate the cDNA clones were constructed by standard

methods using commercially available reagents such as those from Invitrogen, San Diego, CA. The cDNA was primed with oligo dT containing a NotI site, linked with blunt to SalI hemikinased adaptors, cleaved with NotI, sized appropriately by gel electrophoresis, and cloned in a defined orientation into a suitable cloning vector (such as pRKB or pRKD; pRK5B is a precursor of pRK5D that does not contain the SfiI site; see, Holmes et al., Science, 253:1278-1280 (1991)) in the unique XhoI and NotI sites.

DNA sequencing of the clones isolated as described above gave the full-length DNA sequence for PRO300 [herein designated as UNQ263 (DNA40625-1189)] (SEQ ID NO:1) and the derived protein sequence for PRO300.

The entire nucleotide sequence of UNQ263 (DNA40625-1189) is shown in Figure 1 (SEQ ID NO:1). Clone UNQ263 (DNA40625-1189) contains a single open reading frame with an apparent translational initiation site at nucleotide positions 45-47 [Kozak et al., supra] and ending at the stop codon at nucleotide positions 1416-1418 (Figure 1). The predicted polypeptide precursor is 457 amino acids long (Figure 3). Clone UNQ263 (DNA40625-1189) has been deposited with ATCC and is assigned ATCC deposit no. _____.

Analysis of the amino acid sequence of the full-length PRO300 polypeptide suggests that portions of it possess significant homology to the Diff 33 protein. More specifically, an analysis of the Dayhoff database (version 35.45 SwissProt 35) evidenced significant homology between the PRO300 amino acid sequence and the following Dayhoff sequence, HSU49188_1.

EXAMPLE 2: Use of PRO300-encoding DNA as a hybridization probe

The following method describes use of a nucleotide sequence encoding PRO300 as a hybridization probe.

DNA comprising the coding sequence of full-length PRO300 (as shown in Figure 1, SEQ ID NO:1) is employed as a probe to screen for homologous DNAs (such as those encoding naturally-occurring variants of PRO300) in human tissue cDNA libraries or human tissue genomic libraries.

Hybridization and washing of filters containing either library DNAs is performed under the following high stringency conditions. Hybridization of radiolabeled UNQ263 (DNA40625)-derived probe to the filters is performed in a solution of 50% formamide, 5x SSC, 0.1% SDS, 0.1% sodium pyrophosphate, 50 mM sodium phosphate, pH 6.8, 2x Denhardt's solution, and 10% dextran sulfate at 42°C for 20 hours. Washing of the filters is

performed in an aqueous solution of 0.1x SSC and 0.1% SDS at 42°C.

DNAs having a desired sequence identity with the DNA encoding full-length native sequence PRO300 polypeptide can then be identified using standard techniques known in the art.

EXAMPLE 3: Expression of PRO300 Polypeptides in *E. coli*

This example illustrates the preparation of unglycosylated forms of PRO300 polypeptides by recombinant expression in *E. coli*.

The DNA sequence encoding PRO300 (SEQ ID NO:3) is initially amplified using selected PCR primers. The primers should contain restriction enzyme sites which correspond to the restriction enzyme sites on the selected expression vector. A variety of expression vectors may be employed. An example of a suitable vector is pBR322 (derived from *E. coli*; see Bolivar et al., Gene, 2:95 (1977)) which contains genes for ampicillin and tetracycline resistance. The vector is digested with restriction enzyme and dephosphorylated. The PCR amplified sequences are then ligated into the vector. The vector will preferably include sequences which encode for an antibiotic resistance gene, a trp promoter, a polyhis leader (including the first six STII codons, polyhis sequence, and enterokinase cleavage site), the PRO300 coding region, lambda transcriptional terminator, and an argU gene.

The ligation mixture is then used to transform a selected *E. coli* strain using the methods described in Sambrook et al., supra. Transformants are identified by their ability to grow on LB plates and antibiotic resistant colonies are then selected. Plasmid DNA can be isolated and confirmed by restriction analysis and DNA sequencing.

Selected clones can be grown overnight in liquid culture medium such as LB broth supplemented with antibiotics. The overnight culture may subsequently be used to inoculate a larger scale culture. The cells are then grown to a desired optical density, during which the expression promoter is turned on.

After culturing the cells for several more hours, the cells can be harvested by centrifugation. The cell pellet obtained by the centrifugation can be solubilized using various agents known in the art, and the solubilized PRO300 polypeptide can then be purified using a metal chelating column under conditions that allow tight binding of the polypeptide.

EXAMPLE 4: Expression of PRO300 Polypeptides in Mammalian Cells

This example illustrates preparation of glycosylated forms of PRO300 polypeptides by recombinant expression in mammalian cells.

The vector, pRK5 (see EP 307,247, published March 15, 1989), is employed as the expression vector. Optionally, the PRO300-encoding DNA is ligated into pRK5 with selected restriction enzymes to allow insertion of the PRO300-encoding DNA using ligation methods such as described in Sambrook et al., supra. The resulting vector is called pRK5-PRO300.

In one embodiment, the selected host cells may be 293 cells. Human 293 cells (ATCC CCL 1573) are grown to confluence in tissue culture plates in medium such as DMEM supplemented with fetal calf serum and optionally, nutrient components and/or antibiotics. About 10 μg pRK5-PRO300 DNA is mixed with about 1 μg DNA encoding the VA RNA gene [Thimmappaya et al., Cell, 31:543 (1982)] and dissolved in 500 μl of 1 mM Tris-HCl, 0.1 mM EDTA, 0.227 M $CaCl_2$. To this mixture is added, dropwise, 500 μl of 50 mM HEPES (pH 7.35), 280 mM NaCl, 1.5 mM $NaPO_4$, and a precipitate is allowed to form for 10 minutes at 25°C. The precipitate is suspended and added to the 293 cells and allowed to settle for about four hours at 37°C. The culture medium is aspirated off and 2 ml of 20% glycerol in PBS is added for 30 seconds. The 293 cells are then washed with serum free medium, fresh medium is added and the cells are incubated for about 5 days.

Approximately 24 hours after the transfections, the culture medium is removed and replaced with culture medium (alone) or culture medium containing 200 μCi/ml ^{35}S-cysteine and 200 μCi/ml ^{35}S-methionine. After a 12 hour incubation, the conditioned medium is collected, concentrated on a spin filter, and loaded onto a 15% SDS gel. The processed gel may be dried and exposed to film for a selected period of time to reveal the presence of PRO300 polypeptide. The cultures containing transfected cells may undergo further incubation (in serum free medium) and the medium is tested in selected bioassays.

In an alternative technique, PRO300-encoding DNA may be introduced into 293 cells transiently using the dextran sulfate method described by Somparyrac et al., Proc. Natl. Acad. Sci., 12:7575 (1981). 293 cells are grown to maximal density in a spinner flask and 700 μg pRK5-PRO300 DNA is added. The cells are first concentrated from the spinner flask by centrifugation and washed with PBS. The DNA-dextran precipitate is incubated on the cell pellet for four hours. The cells are treated with 20% glycerol for 90 seconds, washed with tissue culture medium, and re-introduced into the spinner flask containing tissue culture

medium, 5 μg/ml bovine insulin and 0.1 μg/ml bovine transferrin. After about four days, the conditioned media is centrifuged and filtered to remove cells and debris. The sample containing expressed PRO300 polypeptide can then be concentrated and purified by any selected method, such as dialysis and/or column chromatography.

In another embodiment, PRO300 polypeptide can be expressed in CHO cells. The pRK5-PRO300 vector can be transfected into CHO cells using known reagents such as $CaPO_4$ or DEAE-dextran. As described above, the cell cultures can be incubated, and the medium replaced with culture medium (alone) or medium containing a radiolabel such as ^{35}S-methionine. After determining the presence of PRO300 polypeptide, the culture medium may be replaced with serum free medium. Preferably, the cultures are incubated for about 6 days, and then the conditioned medium is harvested. The medium containing the expressed PRO300 polypeptide can then be concentrated and purified by any selected method.

Epitope-tagged PRO300 polypeptide may also be expressed in host CHO cells. The PRO300-encoding DNA may be subcloned out of the pRK5 vector. The subclone insert can undergo PCR to fuse in frame with a selected epitope tag such as a poly-his tag into a Baculovirus expression vector. The poly-his tagged PRO300-encoding DNA insert can then be subcloned into a SV40 driven vector containing a selection marker such as DHFR for selection of stable clones. Finally, the CHO cells can be transfected (as described above) with the SV40 driven vector. Labeling may be performed, as described above, to verify expression. The culture medium containing the expressed poly-His tagged PRO300 polypeptide can then be concentrated and purified by any selected method, such as by Ni^{2+}-chelate affinity chromatography.

EXAMPLE 5: Expression of a PRO300 Polypeptide in Yeast

The following method describes recombinant expression of PRO300 polypeptides in yeast.

First, yeast expression vectors are constructed for intracellular production or secretion of PRO300 polypeptide from the ADH2/GAPDH promoter. DNA encoding the PRO300 polypeptide of interest, a selected signal peptide and the promoter is inserted into suitable restriction enzyme sites in the selected plasmid to direct intracellular expression of the PRO300 polypeptide. For secretion, DNA encoding the PRO300 polypeptide can be cloned into the selected plasmid, together with DNA encoding the ADH2/GAPDH promoter, the yeast

alpha-factor secretory signal/leader sequence, and linker sequences (if needed) for expression of the PRO300 polypeptide.

Yeast cells, such as yeast strain AB110, can then be transformed with the expression plasmids described above and cultured in selected fermentation media. The transformed yeast supernatants can be analyzed by precipitation with 10% trichloroacetic acid and separation by SDS-PAGE, followed by staining of the gels with Coomassie Blue stain.

Recombinant PRO300 polypeptide can subsequently be isolated and purified by removing the yeast cells from the fermentation medium by centrifugation and then concentrating the medium using selected cartridge filters. The concentrate containing the PRO300 polypeptide may further be purified using selected column chromatography resins.

EXAMPLE 6: Expression of PRO300 Polypeptides in Baculovirus

The following method describes recombinant expression of PRO300 polypeptides in Baculovirus.

The PRO300-encoding DNA is fused upstream of an epitope tag contained within a baculovirus expression vector. Such epitope tags include poly-his tags and immunoglobulin tags (like Fc regions of IgG). A variety of plasmids may be employed, including plasmids derived from commercially available plasmids such as pVL1393 (Novagen). Briefly, the PRO300-encoding DNA or the desired portion of the PRO300-encoding DNA (such as the sequence encoding the extracellular domain of a transmembrane protein) is amplified by PCR with primers complementary to the 5' and 3' regions. The 5' primer may incorporate flanking (selected) restriction enzyme sites. The product is then digested with those selected restriction enzymes and subcloned into the expression vector.

Recombinant baculovirus is generated by co-transfecting the above plasmid and BaculoGold™ virus DNA (Pharmingen) into *Spodoptera frugiperda* ("Sf9") cells (ATCC CRL 1711) using lipofectin (commercially available from GIBCO-BRL). After 4 to 5 days of incubation at 28°C, the released viruses are harvested and used for further amplifications. Viral infection and protein expression is performed as described by O'Reilley et al., Baculovirus expression vectors: A laboratory Manual, Oxford:Oxford University Press (1994).

Expressed poly-his tagged PRO300 polypeptide can then be purified, for example, by Ni^{2+}-chelate affinity chromatography as follows. Extracts are prepared from recombinant virus-infected Sf9 cells as described by Rupert et al., Nature, 362:175-179 (1993). Briefly,

Sf9 cells are washed, resuspended in sonication buffer (25 mL Hepes, pH 7.9; 12.5 mM $MgCl_2$; 0.1 mM EDTA; 10% Glycerol; 0.1% NP-40; 0.4 M KCl), and sonicated twice for 20 seconds on ice. The sonicates are cleared by centrifugation, and the supernatant is diluted 50-fold in loading buffer (50 mM phosphate, 300 mM NaCl, 10% Glycerol, pH 7.8) and filtered through a 0.45 μm filter. A Ni^{2+}-NTA agarose column (commercially available from Qiagen) is prepared with a bed volume of 5 mL, washed with 25 mL of water and equilibrated with 25 mL of loading buffer. The filtered cell extract is loaded onto the column at 0.5 mL per minute. The column is washed to baseline A_{280} with loading buffer, at which point fraction collection is started. Next, the column is washed with a secondary wash buffer (50 mM phosphate; 300 mM NaCl, 10% Glycerol, pH 6.0), which elutes nonspecifically bound protein. After reaching A_{280} baseline again, the column is developed with a 0 to 500 mM Imidazole gradient in the secondary wash buffer. One mL fractions are collected and analyzed by SDS-PAGE and silver staining or western blot with Ni^{2+}-NTA-conjugated to alkaline phosphatase (Qiagen). Fractions containing the eluted His_{10}-tagged PRO300 polypeptide are pooled and dialyzed against loading buffer.

Alternatively, purification of the IgG tagged (or Fc tagged) PRO300 polypeptide can be performed using known chromatography techniques, including for instance, Protein A or protein G column chromatography.

EXAMPLE 7: Preparation of Antibodies that Bind PRO300 Polypeptides

This example illustrates the preparation of monoclonal antibodies which can specifically bind to PRO300 polypeptides.

Techniques for producing the monoclonal antibodies are known in the art and are described, for instance, in Goding, supra. Immunogens that may be employed include purified PRO300 polypeptide, fusion proteins containing a PRO300 polypeptide, and cells expressing recombinant PRO300 polypeptide on the cell surface. Selection of the immunogen can be made by the skilled artisan without undue experimentation.

Mice, such as Balb/c, are immunized with the PRO300 immunogen emulsified in complete Freund's adjuvant and injected subcutaneously or intraperitoneally in an amount from 1-100 micrograms. Alternatively, the immunogen is emulsified in MPL-TDM adjuvant (Ribi Immunochemical Research, Hamilton, MT) and injected into the animal's hind foot pads. The immunized mice are then boosted 10 to 12 days later with additional immunogen emulsified

in the selected adjuvant. Thereafter, for several weeks, the mice may also be boosted with additional immunization injections. Serum samples may be periodically obtained from the mice by retro-orbital bleeding for testing in ELISA assays to detect anti-PRO300 polypeptide antibodies.

After a suitable antibody titer has been detected, the animals "positive" for antibodies can be injected with a final intravenous injection of PRO300 polypeptide. Three to four days later, the mice are sacrificed and the spleen cells are harvested. The spleen cells are then fused (using 35% polyethylene glycol) to a selected murine myeloma cell line such as P3X63AgU.1, available from ATCC, No. CRL 1597. The fusions generate hybridoma cells which can then be plated in 96 well tissue culture plates containing HAT (hypoxanthine, aminopterin, and thymidine) medium to inhibit proliferation of non-fused cells, myeloma hybrids, and spleen cell hybrids.

The hybridoma cells will be screened in an ELISA for reactivity against PRO300 polypeptide. Determination of "positive" hybridoma cells secreting the desired monoclonal antibodies against a PRO300 polypeptide is within the skill in the art.

The positive hybridoma cells can be injected intraperitoneally into syngeneic Balb/c mice to produce ascites containing the anti-PRO300 polypeptide monoclonal antibodies. Alternatively, the hybridoma cells can be grown in tissue culture flasks or roller bottles. Purification of the monoclonal antibodies produced in the ascites can be accomplished using ammonium sulfate precipitation, followed by gel exclusion chromatography. Alternatively, affinity chromatography based upon binding of antibody to protein A or protein G can be employed.

Deposit of Material

The following materials have been deposited with the American Type Culture Collection, 12301 Parklawn Drive, Rockville, MD, USA (ATCC):

Material	ATCC Dep. No.	Deposit Date
DNA40625-1189		April 21, 1998

This deposit was made under the provisions of the Budapest Treaty on the International Recognition of the Deposit of Microorganisms for the Purpose of Patent Procedure and the Regulations thereunder (Budapest Treaty). This assures maintenance of a viable culture of the

deposit for 30 years from the date of deposit. The deposit will be made available by ATCC under the terms of the Budapest Treaty, and subject to an agreement between Genentech, Inc. and ATCC, which assures permanent and unrestricted availability of the progeny of the culture of the deposit to the public upon issuance of the pertinent U.S. patent or upon laying open to the public of any U.S. or foreign patent application, whichever comes first, and assures availability of the progeny to one determined by the U.S. Commissioner of Patents and Trademarks to be entitled thereto according to 35 USC §122 and the Commissioner's rules pursuant thereto (including 37 CFR §1.14 with particular reference to 886 OG 638).

The assignee of the present application has agreed that if a culture of the materials on deposit should die or be lost or destroyed when cultivated under suitable conditions, the materials will be promptly replaced on notification with another of the same. Availability of the deposited material is not to be construed as a license to practice the invention in contravention of the rights granted under the authority of any government in accordance with its patent laws.

The foregoing written specification is considered to be sufficient to enable one skilled in the art to practice the invention. The present invention is not to be limited in scope by the construct deposited, since the deposited embodiment is intended as a single illustration of certain aspects of the invention and any constructs that are functionally equivalent are within the scope of this invention. The deposit of material herein does not constitute an admission that the written description herein contained is inadequate to enable the practice of any aspect of the invention, including the best mode thereof, nor is it to be construed as limiting the scope of the claims to the specific illustrations that it represents. Indeed, various modifications of the invention in addition to those shown and described herein will become apparent to those skilled in the art from the foregoing description and fall within the scope of the appended claims.

WHAT IS CLAIMED IS:

1. Isolated nucleic acid comprising DNA having at least an 80% sequence identity to (a) a DNA molecule encoding a PRO300 polypeptide comprising the sequence of amino acid residues 1 to 457 of Figure 2 (SEQ ID NO:3), or (b) the complement of the DNA molecule of (a).

2. The nucleic acid of Claim 1 wherein said DNA comprises the nucleotide sequence of SEQ ID NO:1 or its complement.

3. The nucleic acid of Claim 1 wherein said DNA comprises nucleotides 45-1415 of the nucleotide seqeunce of SEQ ID NO:1 (SEQ ID NO:2).

4. Isolated nucleic acid comprising DNA having at least an 80% sequence identity to (a) a DNA molecule encoding the same mature polypeptide encoded by the human protein cDNA in ATCC Deposit No. _____ (DNA40625-1189), or (b) the complement of the DNA molecule of (a).

5. The nucleic acid of Claim 4 which comprises a DNA molecule encoding the same mature polypeptide encoded by the human protein cDNA in ATCC Deposit No. _____ (DNA40625-1189).

6. A vector comprising the nucleic acid of any one of Claims 1 to 5.

7. The vector of Claim 6 operably linked to control sequences recognized by a host cell transformed with the vector.

8. A host cell comprising the vector of Claim 6.

9. The host cell of Claim 8 wherein said cell is a CHO cell.

10. The host cell of Claim 8 wherein said cell is an *E. coli*.

11. The host cell of Claim 8 wherein said cell is a yeast cell.

12. A process for producing a PRO300 polypeptide comprising culturing the host cell of Claim 8 under conditions suitable for expression of said PRO300 polypeptide and recovering said PRO300 polypeptide from the cell culture.

13. Isolated native sequence PRO300 polypeptide comprising amino acid residues 1 to 457 of Figure 2 (SEQ ID NO:3).

14. Isolated PRO300 polypeptide encoded by the cDNA insert of the vector deposited as ATCC Accession No. _____ (DNA40625-1189).

15. A chimeric molecule comprising a PRO300 polypeptide fused to a heterologous amino acid sequence.

16. The chimeric molecule of Claim 15 wherein said heterologous amino acid sequence is an epitope tag sequence.

17. The chimeric molecule of Claim 15 wherein said heterologous amino acid sequence is a Fc region of an immunoglobulin.

18. An antibody which specifically binds to a PRO300 polypeptide.

19. The antibody of Claim 18 wherein said antibody is a monoclonal antibody.

Abstract of the Disclosure

The present invention is directed to novel polypeptides having homology to the human Diff 33 protein and to nucleic acid molecules encoding those polypeptides. Also provided herein are vectors and host cells comprising those nucleic acid sequences, chimeric polypeptide molecules comprising the polypeptides of the present invention fused to heterologous polypeptide sequences, antibodies which bind to the polypeptides of the present invention and to methods for producing the polypeptides of the present invention.

FIGURE 1

GCCCCGCGCCCGGCGCCGGGCGCCCGAAGCCGGGAGCCACCGCC
<MET {trans=1-s, dir=f, res=1}
ATGGGGGCCTGCCTGGGAGCCTGCTCCCTGCTCAGCTGCGCGTCCTGCCTCTGCGGCTCT
GCCCCCTGCATCCTGTGCAGCTGCTGCCCCGCCAGCCGCAACTCCACCGTGAGCCGCCTC
ATCTTCACGTTCTTCCTCTTCCTGGGGGTGCTGGTGTCCATCATTATGCTGAGCCCGGGC
GTGGAGAGTCAGCTCTACAAGCTGCCCTGGGTGTGTGAGGAGGGGGCCGGGATCCCCACC
GTCCTGCAGGGCCACATCGACTGTGGCTCCCTGCTTGGCTACCGCGCTGTCTACCGCATG
TGCTTCGCCACGGCGGCCTTCTTCTTCTTCTTTTTCACCCTGCTCATGCTCTGCGTGAGC
AGCAGCCGGGACCCCCGGGCTGCCATCCAGAATGGGTTTTGGTTCTTTAAGTTCCTGATC
CTGGTGGGCCTCACCGTGGGTGCCTTCTACATCCCTGACGGCTCCTTCACCAACATCTGG
TTCTACTTCGGCGTCGTGGGCTCCTTCCTCTTCATCCTCATCCAGCTGGTGCTGCTCATC
GACTTTGCGCACTCCTGGAACCAGCGGTGGCTGGGCAAGGCCGAGGAGTGCGATTCCCGT
GCCTGGTACGCAGGCCTCTTCTTCTTCACTCTCCTCTTCTACTTGCTGTCGATCGCGGCC
GTGGCGCTGATGTTCATGTACTACACTGAGCCCAGCGGCTGCCACGAGGGCAAGGTCTTC
ATCAGCCTCAACCTCACCTTCTGTGTCTGCGTGTCCATCGCTGCTGTCCTGCCCAAGGTC
CAGGACGCCCAGCCCAACTCGGGTCTGCTGCAGGCCTCGGTCATCACCCTCTACACCATG
TTTGTCACCTGGTCAGCCCTATCCAGTATCCCTGAACAGAAATGCAACCCCCATTTGCCA
ACCCAGCTGGGCAACGAGACAGTTGTGGCAGGCCCCGAGGGCTATGAGACCCAGTGGTGG
GATGCCCCGAGCATTGTGGGCCTCATCATCTTCCTCCTGTGCACCCTCTTCATCAGTCTG
CGCTCCTCAGACCACCGGCAGGTGAACAGCCTGATGCAGACCGAGGAGTGCCCACCTATG
CTAGACGCCACACAGCAGCAGCAGCAGCAGGTGGCAGCCTGTGAGGGCCGGGCCTTTGAC
AACGAGCAGGACGGCGTCACCTACAGCTACTCCTTCTTCCACTTCTGCCTGGTGCTGGCC
TCACTGCACGTCATGATGACGCTCACCAACTGGTACAAGCCCGGTGAGACCCGGAAGATG
ATCAGCACGTGGACCGCCGTGTGGGTGAAGATCTGTGCCAGCTGGGCAGGGCTGCTCCTC
TACCTGTGGACCCTGGTAGCCCCACTCCTCCTGCGCAACCGCGACTTCAGCTGAGGCAGC
CTCACAGCCTGCCATCTGGTGCCTCCTGCCACCTGGTGCCTCTCGGCTCGGTGACAGCCA
ACCTGCCCCCTCCCCACACCAATCAGCCAGGCTGAGCCCCCACCCCTGCCCCAGCTCCAG
GACCTGCCCCTGAGCCGGGCCTTCTAGTCGTAGTGCCTTCAGGGTCCGAGGAGCATCAGG
CTCCTGCAGAGCCCCATCCCCCCGCCACACCCACACGGTGGAGCTGCCTCTTCCTTCCCC
TCCTCCCTGTTGCCCATACTCAGCATCTCGGATGAAAGGGCTCCCTTGTCCTCAGGCTCC
ACGGGAGCGGGGCTGCTGGAGAGAGCGGGGAACTCCCACCACAGTGGGGCATCCGGCACT
GAAGCCCTGGTGTTCCTGGTCACGTCCCCCAGGGGACCCTGCCCCCTTCCTGGACTTCGT
GCCTTACTGAGTCTCTAAGACTTTTTCTAATAAACAAGCCAGTGCGTGTAAAAAAAA

FIGURE 2

MGACLGACSLLSCASCLCGSAPCILCSCCPASRNSTVSRLIFTFFLFLGVLVSIIMLSPG
VESQLYKLPWVCEEGAGIPTVLQGHIDCGSLLGYRAVYRMCFATAAFFFFFFTLLMLCVS
SSRDPRAAIQNGFWFFKFLILVGLTVGAFYIPDGSFTNIWFYFGVVGSFLFILIQLVLLI
DFAHSWNQRWLGKAEECDSRAWYAGLFFFTLLFYLLSIAAVALMFMYYTEPSGCHEGKVF
ISLNLTFCVCVSIAAVLPKVQDAQPNSGLLQASVITLYTMFVTWSALSSIPEQKCNPHLP
TQLGNETVVAGPEGYETQWWDAPSIVGLIIFLLCTLFISLRSSDHRQVNSLMQTEECPPM
LDATQQQQQVAACEGRAFDNEQDGVTYSYSFFHFCLVLASLHVMMTLTNWYKPGETRKM
ISTWTAVWVKICASWAGLLLYLWTLVAPLLLRNRDFS

FIGURE 3A

```
1226511          1 CAGCCGGGACCCCCGGGCTGCCATCCAGAATGGGTTTTGGTTCTTTAAGT
<DNA35930:     1 CAGCCGGGACCCCCGGGCTGCCATCCAGAATGGGTTTTGGTTCTTTAAGT

1226511         51 TCCTGATCCTGGTGGGCCTCACCGTGGGTGCCTTCTACATTCCTGACGGC
1429622          1      ATCCTGGTGGGCCTCACCGTGGGTGCCTTCTACATCCCTGACGGC
<DNA35930:    51 TCCTGATCCTGGTGGGCCTCACCGTGGGTGCCTTCTACATTCCTGACGGC

1226511        101 TCCTTCACCAACATCTGGTTCTACTTCGGCGTCGTGGGCTCCTTCCTCTT
1429622         46 TCCTTCACCAACATCTGGTTCTACTTCGGTGTCGTGGGCTCCTTCCTCTT
1581204          1               TCTGGTTCTACTTCGGTGTCGTGGGCTCCTTCCTCTT
<DNA35930:   101 TCCTTCACCAACATCTGGTTCTACTTCGGTGTCGTGGGCTCCTTCCTCTT

1226511        151 CATCCTCATCCAGCTGGTGCTGCTCATCGACTTTGCGCACTCCTGGAACC
1429622         96 CATCCTCATCCAGCTGGTGCTGCTCATCGACTTTGCGCACTCCTGGAACC
1581204         38 CATCCTCATCCAGCTGGTGCTGCTCATCGACTTTGCGCACTCCTGGAACC
<DNA35930:   151 CATCCTCATCCAGCTGGTGCTGCTCATCGACTTTGCGCACTCCTGGAACC

1226511        201 AGCGGTGGCTGGGCAAGGCCGAGGAGTGCGATTCCC-GTGCCTGGTACGC
1429622        146 AGCGGTGGCTGGGCAAGGCCGAGGAGTGCGATTCCC-GTGCCTGGTACGC
1581204         88 AGCGGTGGCTGGGCAAGGCCGAGGAGTGCGATTCCC-GTGCCTGGNACGC
W49844           1                ATTCGGCAGGAGGGCGATTCCCAGTGCCTGGTACGC
<DNA35930:   201 AGCGGTGGCTGGGCAAGGCCGAGGAGTGCGATTCCCAGTGCCTGGTACGC

1226511        251 AGGCCTCTT
1429622        196 AGGCCTCTTCTTCTTCACTCTCCTCTTCTACTTGCTGT
1581204        138 AGGCCTCTTCTTCTTCACTCTCCTCTTCTACTTGCTGTCGATCGCNGCCG
W49844          37 AGGCCTCTTCTTCTTCACTCTCCTCTTCTACTTGCTGTCGATCGCGGCCG
<DNA35930:   251 AGGCCTCTTCTTCTTCACTCTCCTCTTCTACTTGCTGTCGATCGCGGCCG

1581204        188 TGGCGCTGATGTTCATGTACTACA
W49844          87 TGGCGCTGATGTTCATGTACTACACTGAGCCCAGCGGCTGCCACGAGGGC
<DNA35930:   301 TGGCGCTGATGTTCATGTACTACACTGAGCCCAGCGGCTGCCACGAGGGC
```

FIGURE 3B

```
W49844          137 AAGGTCTTCATCAGCCTCAACCTCACCTTCTGTGTCTGCGTGTCCATCGC
<DNA35930:    351 AAGGTCTTCATCAGCCTCAACCTCACCTTCTGTGTCTGCGTGTCCATCGC

W49844          187 TGCTGTCCTGCCCAAGGTCCAGGACGCCCAGCCCAACTCGGGTCTGCTGC
<DNA35930:    401 TGCTGTCCTGCCCAAGGTCCAGGACGCCCAGCCCAACTCGGGTCTGCTGC

W49844          237 AGGCCTCGGTCATCACCCTCTACACCATGTTTGTCACCTGGTCAGCCCTA
<DNA35930:    451 AGGCCTCGGTCATCACCCTCTACACCATGTTTGTCACCTGGTCAGCCCTA

W49844          287 TCCAAGTATCCCTGAACAAGAAATGCAACCCCCATTTGCCAACCCAGCTG
<DNA35930:    501 TCCAAGTATCCCTGAACAAGAAATGCAACCCCCATTTGCCAACCCAGCTG

W49844          337 GGCAACGAGACAGTTGTGGCAGGCCCGAGGGTATTGAGACCCAGTGGTGG
<DNA35930:    551 GGCAACGAGACAGTTGTGGCAGGCCCGAGGGTATTGAGACCCAGTGGTGG

W49844          387 GATGCCCGAGCATTGTGGGCCTCATCATCTTCCTCCTGTGCACCCTCTTC
<DNA35930:    601 GATGCCCGAGCATTGTGGGCCTCATCATCTTCCTCCTGTGCACCCTCTTC

W49844          437 ATCAGTCTGCGCTCCTCAGGCTCCACGGGACGGGGCTGCTGGAGAGAGCG
<DNA35930:    651 ATCAGTCTGCGCTCCTCAGGCTCCACGGGACGGGGCTGCTGGAGAGAGCG

W49844          487 GGGAACTCCA
<DNA35930:    701 GGGAACTCCA
```

FIGURE 4A

```
1520270          1 GCGCCCCGCGCCCGGCGCCGGGCGCCCGAAGCCGGGAGCCGCCGCCATGG
1662702          1      CCGCGCCCGGCGCCGGGCGCCCGAAGCCGGGAGCCGCCGCCATGG
1929803          1            CCGGCGCCGGGCGCC-GAAGCCGGGAGCCGCCGCCATGG
2520045          1                 GCCGGGCGCCCGAAGCCGGGAGCCGCCGCCATGG
2520349          1                 GCCGGGCGCCCGAAGCCGGGAGCCGCCGCCATGG
2487474          1                        GCCCGAAGCCGGGAGCCGCCGCCATGG
1223450          1                            GAAGCCGGGAGCCGCCGCCATGG
774971           1                                 CGGGAGCCGCCGCCATGG
2741948          1                                 CGGGAGCCGCCGCCATGG
1712126          1                                   GGAGCCGCCGCCATGG
                   ++++++++++++++++++++++++++++++++++++++++++++++++++
<consen01>       1 GCGCCCCGCGCCCGGCGCCGGGCGCCCGAAGCCGGGAGCCGCCGCCATGG

1520270         51 GGGCCTGCCTGGGAGCCTGCTCCCTGCTCAGCTGCGCGTCCTGCCTCTGC
1662702         46 GGGCCTGCCTGGGAGCCTGCTCCCTGCTCAGCTGCGCGTCCTGCCTCTGC
1929803         39 GGGCCTGCCTGGGAGCCTGCTCCCTGCTCAGCTGCGCGTCCTGCCTCTGC
2520045         35 GGGCCTGCCTGGGAGCCTGCTCCCTGCTCAGCTGCGCGTCCTGCCTCTGC
2520349         35 GGGCCTGCCTGGGAGCCTNCTCCCTGCTCAGCTGCGCGTCCTGCCTCTGC
2487474         28 GGGCCTGCCTGGGAGCCTGCTCCCTGCTCAGCTGCGCGTCCTGCCTCTGC
1223450         24 GGGCCTGCCTGGGAGCCTGCTCCCTGCTCAGCTGCGCGTCCTGCCTCTGC
774971          19 GGGCCTGCCTGGGAGCCTGCTCCCTGCTCAGCTGCGCGTCCTGCCTCTGC
2741948         19 GGGCCTGCCTGGGAGCCTGCTCCCTGCTCAGCTGCGCGTCCTGCCTCTGC
1712126         17 GGGCCTGCCTGGGAGCCTGCTCCCTGCTCAGCTGCGCGTCCTGCCTCTGC
                   ++++++++++++++++++.+++++++++++++++++++++++++++++++
<consen01>      51 GGGCCTGCCTGGGAGCCTGCTCCCTGCTCAGCTGCGCGTCCTGCCTCTGC

1520270        101 GGCTCTGCCCCCTGCATCCTGTGCAGCTGCTGCCCCGCCAGCCGCAACTC
1662702         96 GGCTCTGCCCCCTGCATCCTGTGCAGCTGCTGCCCCGCCAGCCGCAACTC
1929803         89 NGCTCTGCCCCCTGCATCATGTGCAGCTGCTGCCCCGCCAGCCGCAACTC
2520045         85 NGCTCTNCCCCCTGCATCCTGTGCAGCTGCTGCCCCGCCAGCCGCAACTC
2520349         85 NGCTCTACCCCCTGCATCCTGTGCAG-TGCTGCCCCGCCAGCCGCAACTC
2487474         78 GGCTCTGCCCCCTGCATCCTGTGCAGCTGCTGCCCCGCCAGCCGCAACTC
1223450         74 GGCTCTGCCCCCTGCATCCTGTGCAGCTGCTGCCCCGCCAGCCGCAACTC
774971          69 GGCTCTGCCCCCTGCATCCTGTGCAGCTGCTGCCCCGNCAGCCGCAACTC
```

FIGURE 4B

```
2741948          69 GGCTCTGCCCCCTGCATCCTGTGCAGCTGCTGCCCCGCCAGCCGCAACTC
1712126          67 NGCTCTGCCCCCTGCATCCTGTGCAGCTGCTGCCCCGCCAGCNGCAACTC
                    .+++++.++++++++++.++++++++++++++++.++++.+++++++
<consen01>      101 GGCTCTGCCCCCTGCATCCTGTGCAGCTGCTGCCCCGCCAGCCGCAACTC

1520270         151 CACCGTGAGCCGCCTCATCTTCACGTTCTTCCTCTTCCTGGGGGTGC
1662702         146 CACCGTGAGCCGCCTCATCTTCACGTTCTTCCTCTTCCTGGGGGTGCTGG
1929803         139 CACCGTGAGCCGCCTCATCTTCACGTTCTTCCTCTTCCTGGGGGTGCTGG
2520045         135 CACCGTGAGCCGCCTCATCTTCACGTTCTTCCTCTTCCTGGGGGTGCTGG
2520349         134 CACCGTGAGCCGCCTCATCTTCACGTTCTTCCTCTTCCTGGGGGTGCTGG
2487474         128 CACCGTGAGCCGCCTCATCTTCACGTTCTTCCTCTTCCTGGGGGTGCTGG
1223450         124 CACCGTGAGCCGCCTCATCTTCACGTTCTTCCTCTTCCTGGGGGTGCTGG
774971          119 CACCGTGAGCCGCCTCATCTTCACGTTCTTCCTCTTCCTGGGGGTGCTGG
2741948         119 CACCGTGAGCCGCCTCATCTTCACGTTCTTCCTCTTCCTGGGGGTGCTGG
1712126         117 CACCGTGAGCCGCCTCATCTTCACGTTCTTCCTCTTCCTGGGGGTGCTGG
1843956           1                     CTTCACGTTCTTCCTCTTCCTGGGGGTGCTGG
                    ++++++++++++++++++++++++++++++++++++++++++++++++++++
<consen01>      151 CACCGTGAGCCGCCTCATCTTCACGTTCTTCCTCTTCCTGGGGGTGCTGG

1662702         196 TGTCCATCATTATGCTGAGCCCGGGCGTGGAGAGTCAGCTCTACAAGCTG
1929803         189 TGTCCATCATTATGCTGAGCCCGGGCGTGGAGAGTCAGCTCTACAAGCTG
2520045         185 TGTCCATCATTATGCTGAGCCCGGGCGTGGAGAGTCAGCTCTACA
2520349         184 TGTCCATCATTATGCTGAGCCCGGGCGTGGAGAGTCAGCTCTACAAGCTG
2487474         178 TGTCCATCATTATGCTGAGCCCGGGCGTGGAGAGTCAGCTCTACAAGCTG
1223450         174 TGTCCATCATTATGCTGAGCCCGGGCGTGGAGAGTCAGCTCTACAAGCTG
774971          169 TGTCCATCATTATGCTGAGCCCGGGCGTGGAGAGTCAGCTCTACAAGCTG
2741948         169 TGTCCATCATTATGCTGAGCCCGGGCGTGGAGAGTCAG-TCTACAAG-TG
1712126         167 TGTCCATCATTATGCTGAGCNCGGGCGTGGAGAGTCAGCTCTACAAGCTG
1843956          33 TGTCCATCATTATGCTGAGCCCGGGCGTGGAGAGTCAGCTCTACAAGCTG
1434590           1  GTCCATCATTATGCTGAGCCCGGGCGTGGAGAGTCAGCTCTACAAGCTG
1434591           1  GTCCATCATTATGCTGAGCCCGGGCGTGGAGAGTCAGCTCTACAAGCTG
                    ++++++++++++++++++++.++++++++++++++++++++++++++++
<consen01>      201 TGTCCATCATTATGCTGAGCCCGGGCGTGGAGAGTCAGCTCTACAAGCTG
```

FIGURE 4C

```
1662702          246 CCC
1929803          239 CCCTGGGTGTGTGANGAGGGGGCCGGGATCCCCACCGTCCT
2520349          234 CCCTGGGTGTGT
2487474          228 CCCTGGGTGTGTGA
1223450          224 CCCTGGGTGTGTGAGGAGGGGGCCGGGATCCCCACCGT
774971           219 CCCTGGG
2741948          217 CCCTGGGTGTGTGAGGAGGGGG
1712126          217 CCC
1843956           83 CCCTGGGTGTGTGAGGAGGGGGCCGGGATCCCCACCGTCCTGCAGGGCCA
1434590           50 CCCTGGGTGTGTGAGGAGGGGGCCGGGATCCCCACCGTCCTGCAGGGCCA
1434591           50 CCCTGGGTGTGTGAGGAGGGGGCCGGGATCCCCACCGTCCTGCAGGGCCA
2121391            1                                                CAGGGCCA
                    +++++++++++++++.++++++++++++++++++++++++++++++++++++
<consen01>       251 CCCTGGGTGTGTGAGGAGGGGGCCGGGATCCCCACCGTCCTGCAGGGCCA

1843956          133 CATCGACTGTGGCTCCCTGCTTGGCTACCGCGCTGTCTACCGCATGTGCT
1434590          100 CATCGACTGTGGCTCCCTGCTTGGCTACCGCGCTGTCTACCGCATGTGCT
1434591          100 CATCGACTGTGGCTCCCTGCTTGGCTACCGCGCTGTCTACCGCATGTGCT
2121391            9 CATCGACTGTGGCTCCCTGCTTGGCTACCGCGCTGTCTACCGCATGTGCT
1705518            1                                              CGCATGTGCT
                    ++++++++++++++++++++++++++++++++++++++++++++++++++++
<consen01>       301 CATCGACTGTGGCTCCCTGCTTGGCTACCGCGCTGTCTACCGCATGTGCT

1843956          183 TCGCCACGGCGGCCTTCTTCTTCTTTTTCACCCTGCTCATGCTCTGCGTG
1434590          150 TCGCCACGGCGGCCTTCTTCTTCTTTTTCACCCTGCTCATGCTCTGCGTG
1434591          150 TCGCCACGGCGGCCTTCTTCTTCTTTTTCACCCTGCTCATGCTCTGCGTG
2121391           59 TCGCCACGGCGGCCTTCTTCTTCTTTTTCACCCTGCTCATGCTCTGCGTG
1705518           11 TCGCCACGGCGGCCTTCTTCTTCTTTTTCACCCTGCTCATGCTCTGCGTG
1226511            1                                          CTCTGCGTG
                    ++++++++++++++++++++++++++++++++++++++++++++++++++++
<consen01>       351 TCGCCACGGCGGCCTTCTTCTTCTTTTTCACCCTGCTCATGCTCTGCGTG
```

FIGURE 4D

```
1843956          233 AANAGCAGCCGGGACCCCCGGGCTGCCATCCAGAATGGGTTTGGGT-CTT
1434590          200 AN-AGCAGCCGGGACCCCCGGGCTGCCATCCAGA
1434591          200 AGCAGCAGCCGGGACCCCCGGGCTGCCATCCAGA
2121391          109 AGCAGCAGCCGGGACCCCCGGGCTGCCATCCAGAATGGGTTTTGGTTCTT
1705518           61 AGCAGCAGCCGGGACCCCCGGGCTGCCATCCAGAATGGGTTTTGGTTCTT
1226511           10 AGCANCAGCCGGGACCCCCGGGCTGCCATCCAGAATGGGTTTTGGTTCTT
DNA35930.init      1       CAGCCGGGACCCCCGGGCTGCCATCCAGAATGGGTTTTGGTTCTT
                     +..+.+++++++++++++++++++++++++++++++++++.+++++++
<consen01>       401 AGCAGCAGCCGGGACCCCCGGGCTGCCATCCAGAATGGGTTTTGGTTCTT

1843956          282 -AAGT-CCTGATTCCTGGT
2121391          159 TAAGTTCCTGAT-CCTGGTGGGCCTCACCGTGGGTGCCTTCTACATCCCT
1705518          111 TAAGTTCCTGAT-CCTGGTGGGCCTCACCGTGGGTGCCTTCTACATTCCT
1226511           60 TAAGTTCCTGAT-CCTGGTGGGCCTCACCGTGGGTGCCTTCTACATTCCT
DNA35930.init     46 TAAGTTCCTGAT-CCTGGTGGGCCTCACCGTGGGTGCCTTCTACATTCCT
1429622            1            AT-CCTGGTGGGCCTCACCGTGGGTGCCTTCTACATCCCT
                     ++++++++++++ ++++++++++++++++++++++++++++++++.+++
<consen01>       451 TAAGTTCCTGAT CCTGGTGGGCCTCACCGTGGGTGCCTTCTACATTCCT

2121391          208 GACGGCTCCTTCACCAACATCTG
1705518          160 GACGGCTCCTTCACCAACATCTGGTTCTACTTCGGCGTCGTGGGCTCCTT
1226511          109 GACGGCTCCTTCACCAACATCTGGTTCTACTTCGGCGTCGTGGGCTCCTT
DNA35930.init     95 GACGGCTCCTTCACCAACATCTGGTTCTACTTCGGTGTCGTGGGCTCCTT
1429622           40 GACGGCTCCTTCACCAACATCTGGTTCTACTTCGGTGTCGTGGGCTCCTT
1581204            1                        TCTGGTTCTACTTCGGTGTCGTGGGCTCCTT
2159726            1                                  TCGGCGTCGTGGGCTCCTT
1815012            1                                            TGGGCTCCTT
HSU47700           1                                              GGCTCCTT
                     ++++++++++++++++++++++++++++++++++.++++++++++++++
<consen01>       500 GACGGCTCCTTCACCAACATCTGGTTCTACTTCGGTGTCGTGGGCTCCTT

1705518          210 CCNCCTTCA
1226511          159 CCTCTTCATCCTCATCCAGCTGGTGCTGCTCATCGACTTTGCGCACTCCT
DNA35930.init    145 CCTCTTCATCCTCATCCAGCTGGTGCTGCTCATCGACTTTGCGCACTCCT
1429622           90 CCTCTTCATCCTCATCCAGCTGGTGCTGCTCATCGACTTTGCGCACTCCT
```

FIGURE 4E

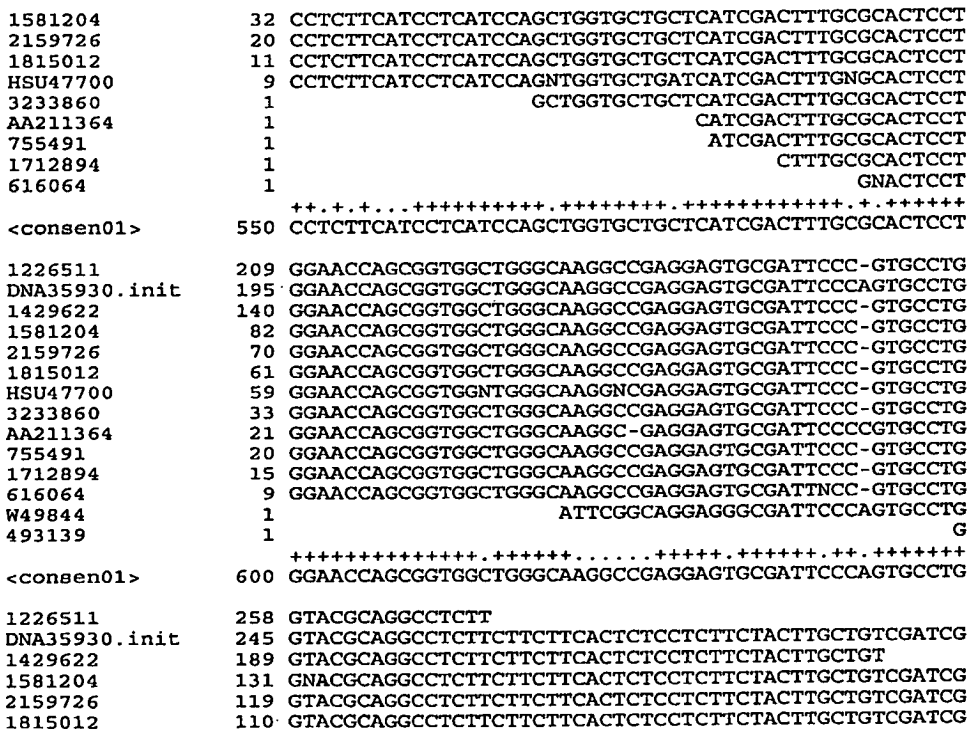

FIGURE 4F

```
HSU47700        108 GTACGCAGGCCTCTTCTTCTTCACTCTCCTCTTCTACTNGATGTCGATCG
3233860          82 GTACGCAGGCCTCTTCTTCTTCACTCTCCTCTTCTACTTGCTGTCGATCG
AA211364         70 GTACGCAGGCCTCTTCTTCTTC
755491           69 GTACGCAGGCCTCTTCTTCTTCACTCTCCTCTTCTACTTGCTGTCGATCG
1712894          64 GTACGCAGGCCTCTTCTTCTTCACTCTCCTCTTCTACTTGCTGTCGATCG
616064           58 GTACGGAGGCCTCTTCTTCTTCANTCTNCTCTTCTACTTGCTGTCGATCG
W49844           31 GTACGCAGGCCTCTTCTTCTTCACTCTCCTCTTCTACTTGCTGTCGATCG
493139            2 GTACGNAGGNCTCTNCTNCTANANTNTCCTNTTCTACTTGCTGTCGATCG
1665446           1                                     ACTTGCTGTCGATCG
R72007            1                                               GATCG
R72471            1                                                ATCG
1508356           1                                                   G
                    +.+++.+++.++++.++.++..+.+.+.++.+++++++.+.+++++++++
<consen01>      650 GTACGCAGGCCTCTTCTTCTTCACTCTCCTCTTCTACTTGCTGTCGATCG

DNA35930.init   295 CGGCCGTGGCGCT-GA-TGTTCATGTACTACACT-GAGCCCAGCGGCTGC
1581204         181 CNGCCGTGGCGCT-GA-TGTTCATGTACTACA
2159726         169 CGGCCGTGGCGCT-GA-TGTTCATGTACTACACT-GAGCCCAGCGGCTGC
1815012         160 CGGCCGTGGCGCT-GA-TGTTCATGTACTACACT-GAGCCCAGCGGCTGC
HSU47700        158 CGGCCGTGGTGCT-GA-TGTNCATGTACTACACT-GAGCCCAGAGGNTGN
3233860         132 CNGCCGTGGCGCT-GA-TGTTCATGTACTACACT-GAGCC-AGCGGCTGC
755491          119 CGGCCGTGGCGCT-GA-TGTTCATGTACTACANT-GAGCCCAGCGGCTGC
1712894         114 CGGCCGTGGCGCT-GA-TGTTCATGTACTACACT-GAGCCCAGCGGCTGC
616064          108 CGGCCGTGGCGNT-GA-TGTTCATGTACTACANT-GAGNCCAGCGGCTGC
W49844           81 CGGCCGTGGCGCT-GA-TGTTCATGTACTACACT-GAGCCCAGCGGCTGC
493139           52 CGGCCGTGGCGCT-GA-TGTTCATGTACTACACT-GAGCCCAGNGGCTGC
1665446          16 CGGCCGTGGCGCT-GA-TGTTCATGTACTACACT-GAGCCCAGCGGCTGC
R72007            6 CGGCCGTGGC-CTCGAATGTTCATGTACTACACT-GAGCCCAGCG-CTGC
R72471            5 CGGCCNTGGC-CT-GAATGTTCATGTACTACANTTGAGCCCAGCG-CTGC
1508356           2 CGGCCGTGGCGCT-GA-TGTTCATGTACTACACT-GAGCCCAGCGGCTGC
876679            1   GCCGNGGCGCT-GA-TGTTCATGTACTACACT-GAGCCCAGCGGCTGC
820198            1      GTGGCGCT-GA-TGTTCATGTACTACACT-GAGCCCAGCGGCTGC
3120012           1                                              GCGGCTGC
1424153           1                                                  CTGC
                    +.+++..++.+.+ ++ +++.+++++++++++.+ +++.++++.++.++.
<consen01>      700 CGGCCGTGGCGCT GA TGTTCATGTACTACACT GAGCCCAGCGGCTGC
```

FIGURE 4G

```
DNA35930.init     342 CACGAGGGCAAGGTCTTCATCAGCCTC-AACCTC-ACCTTCTGTGTCTGC
2159726           216 CACGAGGGCAAG
1815012           207 CACGAGGGCAAGGTCTTCATCAGCCTC-AACCTC-ACCTTNTGTGTCTGC
HSU47700          205 CACGAGGGCAAGGTCTTCATCAGNCTC-AACCTC-ACCTNCTGTGTCTGN
3233860           178 NACGAGGGCAAGGTCTTAATCAGCTTA-ACCTAA-CTTCGGGGTNGNGGG
755491            166 CACGAGGGCAAGGTCTTCATCAGCCTC-AACCTC-ANCTTCTGTGTCTGC
1712894           161 CACGAGGGCAAGGTCTTCATCAGCNTC-AACCTC-ACCTTCTGTGTCTGC
616064            155 CACGAGGGCAAGGTCTTNATCAGNCTT-AANCTT-AACTTCTGTGTCTGC
W49844            128 CACGAGGGCAAGGTCTTCATCAGCCTC-AACCTC-ACCTTCTGTGTCTGC
493139             99 CACGAGGGCAAGGNNTNCATNAGACTC-AACCTN-ACCTANTGTGAATGC
1665446            63 CACGAGGGCAAGGTCTTCATCAGCCTC-AACCTC-ACCTTCTGTGTCTGC
R72007             53 CACGAGGGCAAGGTCTTCATCAGCCTC-AACCTC-ACCTTCTGTGTCTGC
R72471             52 CACGAGGGCAAGGTCTTCATCAGCCTC-AACCTN-ANCNTCTGTGTCTGC
1508356            49 CACGAGGGCAAGGTCTTCATCAGCCTC-AACCTC-ACCTTCTGTGTCTGC
876679             46 CACGAGGGCAAGGTCTTCATCAGCCTN-AACCTC-ACCTTCTGTGTCTGC
820198             43 CACGAGGGCAAGGTCTTNATCAGCCTC-AANCTC-ACCTTCTGTGTCTGC
3120012             9 CACGAGGGCAAGGTCTTCATCAGCCTC-AACCTC-ACCTTCTGTGTCTGC
1424153             5 CACGAGGGCAAGGTCTTCATCAGCCTCCAACCTCCACCTTCTGTGTCTGC
1431207             1  CGGCTCGAGGTGTGTGTGCATGTGAG-TGGTAT-GTGGGTATGTGTGAG
2117449             1                                  CTC-ACCTTCTGTGTCTGC
1282436             1                                      TCTGTGTCTGC
550405              1                                          CCCA
                      ...+...+....+..+......+.... ...... ..............
<consen01>        747 CACGAGGGCAAGGTCTTCATCAGCCTC AACCTC ACCTTCTGTGTCTGC

DNA35930.init     390 GTGTCCATCGCTGCTGTCCTGCCCAAGGTCCAGGACGCCCAGCCCAACTC
1815012           255 GTGTCCATCG-TGCTGTCCT
HSU47700          253 GTGTCCATCGCTGGTGTCCTGCCCAAGGTNCAGGACGNCCAGGCCAACTC
3233860           226 CNANNGNGNGNCCNCCC
755491            214 GTGTCCATCGCTGCTGTCCTGCCCAAGGTCCAGGAAG
1712894           209 GT
616064            203 GTGTTCATCG
W49844            176 GTGTCCATCGCTGCTGTCCTGCCCAAGGTCCAGGACGCCCAGCCCAACTC
493139            147 GTGTCCATCGCTGCNGTCCTGCCCAAAGTCCAGGANGNCCAGNCCACATN
1665446           111 GTGTCCATCGCTGCTGTCCTGCCCAAGGTCCAGGACGCCCAGCCCAACTC
```

FIGURE 4H

```
R72007          101 GTGTCCATCGCTGCTGTCCTGCCCAAGGTCCAGGACGCCCAGCCCAACTC
R72471          100 GTGTCCATCGCTGCTGTCCTGCCCAAGGTCCAGGACGCCCAGCCCAACTC
1508356          97 GTGTCCATCGCTGCTGTCCTGCCCAAGGTCCAGGACGCCCAGCCCAACTC
876679           94 GTGTCCATCGCTGCTGTNCTGCCCAAGGTCCAGGANGACCAGCCCAACTC
820198           91 GTGTCCATCGGTGCTGTCCTGCCCAAGGTCCAGGACGGCCAGCCCAACTG
3120012          57 GTGTCCATCGCTGCTGTCCTGCCCAAGGTCCAGGACGCCCAGCCCAACTC
1424153          55 GTGTCCATCGCTGCTGTCCTGCCCAAGGTCCAGGACGCCCAGCCCAACTC
1431207          48 GGTGTGTGTGTTTGAGTCTGTGGTGTAGCATATGTGGTTTTTGCAAGTTA
2117449          19 GTGTCCATCGCTGCTGTCCTGCCCAAGGTCCAGGACGCCCAGCCCAACTC
1282436          12 GTGTCCATCGCTGCTGTCCTGCCCAAGGTCCAGGACGCCCAGCCCAACTC
550405            5 CGCGTCCGCGCTGCTGTNCTGCCCAAGGTCCAGGACGCCCAGCCCAACTN
1312500           1   GTCCATCGCTGCTGTCCTGCCCAAGGTCCAGGACGCCCAGCCCAACTC
2377129           1  GGCAGGGGTCTTACCAGGGGTCCTCTTGCCAGGACGCCCAGCCCAACTC
274786            1                                           CAGAGCTC
                    ..............................+.+..+........+..+.
<consen01>      795 GTGTCCATCGCTGCTGTCCTGCCCAAGGTCCAGGACGCCCAGCCCAACTC

DNA35930.init   440 GGGTC-T-GCTGCAGGCCTCGGTCATCACCCTCTACACCATGTTTGTCAC
HSU47700        303 GGGTC-TTGTTGGAGGGCTCGGACATCACCCTCTACAACATGTTTGGCAC
W49844          226 GGGTC-T-GCTGCAGGCCTCGGTCATCACCCTCTACACCATGTTTGTCAC
493139          197 GGGTAAT-GCTGCAGGNCTAGGTCATCANCNNATACANCATGTNT
1665446         161 GGGTC-T-GCTGCAGGCCTCGGTCATCACCCTCTACACCATGTTTGTCAC
R72007          151 GGGTC-T-GCTGCAGGCCTCGGTCATCACCCTCTACACCATGTTTGTCAC
R72471          150 GGGTC-T-GCTNCAGGCCTNGGTCATCACCCTCTACACCATGTTTGTCAC
1508356         147 GGGTC-T-GCTGCAGGCCTCGGTCATCACCCTCT
876679          144 GGGTC-T-GCTGCAGGCCTCGGTCATCACCCTATACACCATNTTTGTCAC
820198          141 GGGTC-T-GCTGCAGGNCTCGGTCATCANCCTCTACANCATGTTTGTCAN
3120012         107 GGGTC-T-GCTGCAGGCCTCGGTCATCACCCTCTACACCATGTTTGTCAC
1424153         105 GGGTC-T-GCTGCAGGCCTCGGTCATCACCCTCTACACCATGTTTGTCAC
1431207          98 GGNGA-G-ANGGGCTCAGGTACGTGTGTAACCCGGGGAAGCAGGTGATGG
2117449          69 GGGTC-T-GCTGCAAGCCTCGGTCATCACCCTCTACACCATGTTTGTCAC
1282436          62 GGGTC-T-GCTGCAGGCCTCGGTCATCACCCTCTACACCATGTTTGTCAC
550405           55 GGGTC-T-GCTGCAGGCCTCGGTCATNACCCTNTACACCATGTTTGTNAC
1312500          49 GGGTC-T-GCTGCAGGCCTCGGTCATCACCCTCTACACCATGTTTGTCAC
2377129          50 GGGTC-T-GCTGCAGGCCTCGGTCATCACCCTCTACACCATGTTTGTCAC
```

FIGURE 4I

```
274786              9 GGGTC-T-GCTGCAGGCCTCGGTCATNACCCTCTACACCATGTTTTTAAN
AA130551            1    TC-T-GCTGCAGGCCTCGGTCATCACCCTCTACACCATGTTTGTCAC
AA481928.RC         1 GCAGG-A-GGCACCAGATGGCAGGCTGTGAGGCTGCCTCAGCTGAAGTCG
023897              1              CAGGCCTCGGNCATCACCCTCTACACCATGTTTGTAAC
1453117             1              CAGGCCTCGGTCATCACCCTCTACACCATGTTTGTCAC
1550018             1                        ATCACCCTCTACACCATGTTTGTCAC
1731292             1                             CCTCTACACCATGTTTGTCAC
2417559             1                                  ACACCATGTTTGTCAC
AA410947.RC         1                                        GCTGAAGTCG
1754776             1                                         GTTTGTCAC
                      +.... . .................+.......................
<consen01>        845 GGGTC T GCTGCAGGCCTCGGTCATCACCCTCTACACCATGTTTGTCAC

DNA35930.init     488 CTGGTCAGCCCTATCCAAGTATCCCTGAACAAGAAATGCAACCCCCCATTT
HSU47700          352 CTGGNCAGACCTATCCAN-TATCCCTGAACA-GAAATGTAACCCCCCATTT
W49844            274 CTGGTCAGCCCTATCCAAGTATCCCTGAACAAGAAATGCAACCCCCCATTT
1665446           209 CTGGTCAGCCCTATCCA-GTATCCCTGAACA-GAAATGCNACC
R72007            199 CTGGTCAGCCCTATCCA-GTATCCCTGAACA-GAAATGCAACCCCCCATTT
R72471            198 CTGGTCAGCCCTATCCA-GTATCCCTGAACA-GAAATGCAACCCCCCATTT
876679            192 CTGGTCAGCCCTATCCA-GTATCCCTGAACA-GAAATNCAACCCACNTTT
820198            189 CTGGTCAGCCCTATNCA-GTATCCCTGAACA-GNAATGNAACCNNCATTT
3120012           155 CTGGTCAGCCCTATCCA-GTATCCCTGAACA-GAAATGCAACCCCCCATTT
1424153           153 CTGGTCAGCCCTATCCA-GTATCCCTGAACA-GAAATGCAACCCCCCATTT
1431207           146 AGGGTCAGCCCTATCCA-GTATCCCTGAACA-GAAATGCAACCCCCCATTT
2117449           117 CTGGTCAGCCCTATCCA-GTATCCCTGAACA-GAAATGCAACCCCCCATTT
1282436           110 CTGGTCAGCCCTATCCA-GTATCCCTGAACA-GAAATGCAACCCCCCATTT
550405            103 CTGGTCAGCCCTATCCA-GTATCCCTGAACA-GAAATGCAACCCCCCATTT
1312500            97 CTGGTCAGCCCTATCCA-GTATCCCTGAACA-GAAATGCAACCCCCCATTT
2377129            98 CTGGTCAGCCCTATCCA-GTATCCCTGGTAAGTATGGGCCCAGGCTCAAG
274786             57 CTGGTCAGCCCTATCCA-GTATCCCTGAACA-GAAATGCAANCCCCCATTT
AA130551           46 CTGGTCAGCCCTATCCA-GTATCCCTGAACA-GAAATGCAACCCCCCATTT
AA481928.RC        49 CGGTTGCGCAGGAGGAGTGGGGCTACCAGGGTCCACAGGTAGAGGAGCAG
023897             39 CTGGNCAGCCCTATCCA-GTATCCCTGAACA-GNAATGCAACCCCNCATTT
1453117            39 CTGGTCAGCCCTATCCA-GTATCCCTGAACA-GAAATGCAACCCCCCATTT
1550018            27 CTGGTCAGCCCTATCCA-GTATCCCTGAACA-GAAATGCAACCCCCCATTT
```

FIGURE 4J

```
1731292          22 CTGGTCAGCCCTATCCA-GTATCCCTGAACA-GAAATGCAACCCCCATTT
2417559          17 CTGGTCAGCCCTATCCA-GTATCCCTGAACA-GAAATGCAACCCCCATTT
AA410947.RC      11 CGTTGCTGCAGGAGGAGTGGGGCTACCAGGGTCCACAGGTAGAGGAGCAG
1754776          10 CTGGTCAGCCCTATCCA-GTATCCCTGAACA-GAAATGCAACCCCCATTT
2279337           1   GGTCAGCCCTATCCA-GTATCCCTGAACA-GAAATGCAACCCCCATTT
941498            1   GGTCAGCCCTATCCA-GTATCCCTGAACA-GAAATGCAACCCCCATTT
755392            1        GCCCTATCCA-GTATCCCTGAACA-GAAATGCAACCCCCATTT
548789            1                CCAGTATCCCTGNAACAGAAATGCAACCCCCATTT
3039067           1                   GTATCCCTGAACA-GAAATGCAACCCCCATTT
2376768           1                           GAACA-GAAATGCAACCCCCATTT
1682539           1                                      GCAACCCCCATTT
1804118           1                                         CGGCTCGAGT
                    .......+....+.....+...+..........................
<consen01>      893 CTGGTCAGCCCTATCCAAGTATCCCTGAACAAGAAATGCAACCCCCATTT

DNA35930.init   538 GCCAACCCAGCTGGG-CAACGAGACAGTTGT-GG-CAGGCCC-GAGGG-T
HSU47700        400 GGCAACNCAGNTGNG-AAACGANA-AGTTGT-GG
W49844          324 GCCAACCCAGCTGGG-CAACGAGACAGTTGT-GG-CAGGCCC-GAGGG-T
R72007          247 GCCAACCCAGCTGAG-CAACGAGACAGTTGT-GG-CAGGNCCCGAGGNTG
R72471          246 GCCAACCCAGCTGAG-CAACGAGACAGTTGT-GG-CAGGCCCCGAGGNTG
876679          240 GCCAACCCAGCTGGG-CAACGNGACAGTTC
820198          237 GCCAAACCAGCTGGG
3120012         203 GCCAACCCAGCTGGG-CAACGAGACAGTTGT-GG-CAGGCCCCGAGGGCT
1424153         201 GCCAACCCAGCTGGG-CAACGAGACAGTTGT-GG-CAGGCCCCGAGGG
1431207         194 GCCAACCCAGCTGGG-CAACGA
2117449         165 GCCAACCCAGCTGGG-CAACGAGACAGTTGT-GG-CAGGCCCCGAGGGCT
1282436         158 GCCAACCCAGCTGGG-CAACGAGACAGTTGT-GG-CAGGCCCCGAGGGCT
550405          151 GCCAACCCAGCTGGG-CAACGAGACAGTTGT-GG-CAGGCCCCGAGGGCT
1312500         145 GCCAACCCAGCTGGG-CAACGAGACAGTTGT-GG-CAGGCCCCGAGGGCT
2377129         147 GAGGCCTGGCCTCGT-TCCTGGGTCAGTATC-AG-TCTACTGTGGGGCTG
274786          105 TCCAATNCAGNTGGGNCAANGAGACANTTTT-GG-CAGGCCCCGAGGGCT
AA130551         94 GCCAACCCAGCTGGG-CAACGAGACAGTTGT-GG-CAGGCCCCGAGGGGT
AA481928.RC      99 CCCTACCCAGCTGGG-CAACGAGACAGTTGT-GG-CAGCGCCCGAGGGCT
023897           87 GNCAACCNAGNTGGG-CAACGNGACAGTTNT-GG-CAGGCCCNGAGGGCT
1453117          87 GCCAACCCAGCTGGG-CAACGAGACAGTTGT-GG-CAGGCCCCGAGGGCT
```

FIGURE 4K

```
1550018         75 GCCAACCCAGCTGGG-CAACGAGACAGTTGT-GG-CAGGCCCCGAGGGCT
1731292         70 GCCAACCCAGCTGGG-CAACGAGACAGTTGT-GG-CAGGCCCCGAGGGCT
2417559         65 GCCAACCCAGCTGGG-CAACGAGACAGTTGT-GG-CAGGCCCCGAGGGCT
AA410947.RC     61 CCCTACCCAGCTGGG-CAACGAGACAGTTGTAGG-CACGCCCCGAGGGCT
1754776         58 GCCAACCCAGCTGGG-CAACGAGACAGTTGT-GG-CAGGCCCCGAGGGCT
2279337         47 GCCAACCCAGCTGGG-CAACGAGACAGTTGT-GG-CAGGCCCCGAGGGCT
941498          47 GCCAACCCAGCTGAG-CAACGAGACAGTTGT-GG-CAGGCCCCGAGGGCT
755392          42 GCCAACCCAGCTGAG-CAACGAGACAGTTGT-GG-CAGGCCCCGAGGGCT
548789          36 GCCAACCCAGCTGGG-CAACGAGACAGTTGT-GG-CAGGCCCCGAGGGCT
3039067         32 GCCAACCCAGCTGAG-CAACGAGACAGTTGT-GG-CAGGCCCCGAGGGCT
2376768         24 GCCAACCCAGCTGGG-CAACGAGACAGTTGT-GG-CAGGCCCCGAGGGCT
1682539         14 GCCAACCCAGCTGGG-CAACGAGACAGTTGT-GG-CAGGCCCCGAGGGCT
1804118         11 GCCAACCCAGCTGGG-CAACGAGACAGTTGT-GG-CAGGCCCCGAGGGCT
2193380          1     CCCAGCTGGG-CAACGAGACAGTTGT-GG-CAGGCCCCGAGGGCT
3210913          1     CCCAGCTGGG-CAACGAGACAGTTGT-GGCAGGCCCCGAGGGCT
2321125          1                   CGAGACAGTTGT-GG-CAGGCCCCGAGGGCT
2743359          1                         GTTGT-GG-CAGGCCCCGAGGGCT
1755244          1                         GTTGT-GG-CAGGCCCCGAGGGCT
1752950          1                         GTTGT-GG-CAGGCCCCGAGGGCT
3255760          1                            GT-GG-CAGGCCCCGAGGGCT
1655567          1                                      CCCGAGGGCT
AA477170.RC      1                                      CCCGAGGGCT
                   ...........+... ....+...++.+... .+ ........+.++...
<consen01>     943 GCCAACCCAGCTGGG CAACGAGACAGTTGT GG CAGGCCCCGAGGGCT

DNA35930.init  583 A-TTG-AGACCCAGTGGTGGG-ATGCCC-GAGCATTGTGGGCCTCATCAT
W49844         369 A-TTG-AGACCCAGTGGTGGG-ATGCCC-GAGCATTGTGGGCCTCATCAT
R72007         294 A-T-GGAGACCCAGTTGTGGGGATGCCCCGAGCATTGTNGGCC
R72471         293 A-T-G-AGACCCAGTGGTNGG-ATGCCCCGAGCATTGTGGGCCTCATCAT
3120012        250 A-T-G-AGACCCAGTGGTGGG-ATGCCC-GAGCAT
2117449        212 A-T-G-AGACCCAGTGGTGGG-ATGCCCCGAGCATTGTGGGGCTCATCAT
1282436        205 A-T-G-AGACCCAGTGGTGGG-ATGCCCCGAGCATTGTGGGCCTCATCAT
550405         198 A-T-G-AGACCCAGTGGTGGG-ATGCCCCGAGCATTGTGGGCCTNATNAT
1312500        192 A-T-G-AGACCCAGTGGTGGG-ATGCCCCGAGCATTGTGGGCCTCATCAT
2377129        194 G-G-G-ACCCTCACAGGTGAC-AGGGACATCCCTGCTCCAGCCCA
```

FIGURE 4L

```
274786           153 A-T-N-NGACCCANTGGTGGG-NTGCCCCGAGCATTNTGGGCCTCATNAT
AA130551         141 ANT-G-AGACCCAGTGGTGGG-ATGCCCCGAGCATTGTGGGCCTCATCAT
AA481928.RC      146 A-T-G-AGACCCAGTGGTGGG-ATGCCCCGAGCATTGTGGGCCTCATCAT
023897           134 A-T-G-AGACCCAGTGGTGGG-ATGCCCCGAGCATTGTGGGCCTCATNAT
1453117          134 A-T-G-AGACCCAGTGGTGGG-ATGCCCCGAGCATTGTGGGCCTCATCAT
1550018          122 A-T-G-AGACCCAGTGGTGGG-ATGCCCCGAGCATTGTGGGCCTCATCAT
1731292          117 A-T-G-AGACCCAGTGGTGGG-ATGCCCCGAGCATTGTGGGCCTCATCAT
2417559          112 A-T-G-AGACCCAGTGGTGGG-ATGCCCCGAGCATTGTGGGCCTCATCAT
AA410947.RC      109 A-T-G-AGACCCAGTGGTGG--ATGCCCCGAGCATTGTGGGCCTCATCAT
1754776          105 A-T-G-AGACCCAGTGGTGGG-ATGCCCCGAGCATTGTGGGCCTCATCAT
2279337           94 A-T-G-AGACCCAGTGGTGGG-ATGCCCCGAGCATTGTGGGCCTCATCAT
941498            94 A-T-G-AGACCCAGTGGTGGG-ATGCCCCGAGCATTGTGGGCCTCATCAT
755392            89 A-T-G-AGACCCAGTGGTGGG-ATGCCCCGAGCATTGTGGGCCTCATCAT
548789            83 A-T-G-AGACCCAGTGGTGGG-ATGCCCCGAGCATTGTGGGCCTCATCAT
3039067           79 A-T-G-AGACCCAGTGGTGGG-ATGCCCCGAGCATTGTGGGCCTCATCAT
2376768           71 A-T-G-AGACCCAGTGGTGGG-ATGCCCCGAGCATTGTGGGCCTCATCAT
1682539           61 A-T-G-AGACCCAGTGGTGGG-ATGCCCCGAGCATTGTGGGCCTCATCAT
1804118           58 A-T-G-AGACCCAGTGGTGGG-ATGCCCCGAGCATTGTGGGCCTCATCAT
2193380           43 A-T-G-AGACCCAGTGGTGGG-ATGCCCCGAGCATTGTGGGCCTCATCAT
3210913           44 A-T-G-AGACCCAGTGGTGGG-ATGCCCCGAGCATTGTGGGCCTCATCAT
2321125           30 A-T-G-AGACCCAGTGGTGGG-ATGCCCCGAGCATTGTGGGCCTCATCAT
2743359           23 A-T-G-AGACCCAGTGGTGGG-ATGCCCCGAGCATTGTGGGCCTCATCAT
1755244           23 A-T-G-AGACCCAGTGGTGGG-ATGCCCCGAGCATTGTGGGCCTCATCAT
1752950           23 A-T-G-AGACCCAGTGGTGGG-ATGCCCCGAGCATTGTGGGCCTCATCAT
3255760           20 A-T-G-AGACCCAGTGGTGGG-ATGCCCCGAGCATTGTGGGCCTCATCAT
1655567           11 A-T-G-AGACCCAGTGGTGGG-ATGCCCCGAGCATTGTGGGCCTCATCAT
AA477170.RC       11 A-T-G-AGACCCAGTGGTGGG-ATGCCCCGAGCATTGTGGGCCTCATCAT
H02520             1     G-AGACCCAGTGGTGGG-ATGCCCCGAGCATTGTGGGCCTCATCAT
2623580            1                  GTGGG-ATGCCCCGAGCATTGTGGGCCTCATCAT
732580             1                            GCCCCGAGCATTGTGGGCCTCATCAT
1579688            1                      G-ATGCCCCGAGCATTGTGGGCCTCATCAT
1579888            1                      G-ATGCCCCGAGCATTGTGGGCCTCATCAT
794186             1                                        GTGGGCCTCATCAT
1733505            1                                            GCCTCATCAT
                     . . . ...+.++...++... ..+..+....+.......+.+..++.++
<consen01>       990 A T G AGACCCAGTGGTGGG ATGCCCCGAGCATTGTGGGCCTCATCAT
```

FIGURE 4M

DNA35930.init	629	-CTTCCTCCTGTGCACCCTCTTCATCAGTCTGCGCTCCTCAGGCTCCACG
W49844	415	-CTTCCTCCTGTGCACCCTCTTCATCAGTCTGCGCTCCTCAGGCTCCACG
R72471	339	-CTTCCTNCTGTGCAACCTCTTTAT
2117449	258	-CTTCCTCCTGTGAACC-TCTTCATCAGTCTGCGCTCCTCA
1282436	251	-CTTCCTCCTG
550405	244	-GTTGCTGCTGTGCACCCTNTTNATNA
1312500	238	-CTTCCTCCTGTGCACCCTCTTCATCAGTCTGCGCTCCTC
274786	199	-NTTCCTCCTGTGCACNCTCTTNATNAGTCTGCGNTCCTCAGACCACCGG
AA130551	188	-CTTCCTCCTGTGCACCCTCTTCATCAGTCTGCGCTCCTCAGACCACCGG
AA481928.RC	192	-CTTCCTCCTGTGCACCCTCTTCATCAGTCTGCGCTCCTCAGACCACCGG
023897	180	TNTTCCTCCTNTGTACNCNATTNATTAGTCTNCGTTTCTNANNNCACCGN
1453117	180	-CTTCCTCCTGTGCACCCTCTTCATCAGTCTGCGCTCCTCAGACCACCGG
1550018	168	-CTTCCTCCTGTGCACCCTCTTCATCAGTCTGCG
1731292	163	-CTTCCTCCTGTGCACCCTCTTCATCAGTCTGCGCTCCTCAGACCACCGG
2417559	158	-CTTCCTCCTGTGCACCCTCTTCATCAGTCTGCGCTCCTCAGACCACCGG
AA410947.RC	154	-CTTCCTCCTGTGCACCCTCTTCATCAGTCTGCGCTCCTCAGACCACCGG
1754776	151	-CTTCCTCCTGTGCACCCTCTTCATCAGTCTGCGCTCCTCAGACCACCGG
2279337	140	-CTTCCTCCTGTGCACCCTCTTCATCAGTCTGCGCTCCTCAGACCACCGG
941498	140	-CTTCCTCCTGTGCACCCTCTTCATCAGTCTGCGCTCCTCAGACCACCGG
755392	135	-CTTCCTCCTGTGCACCCTNTTCATCAGTCTGCGNTCCTCAGACCACCGG
548789	129	-CTTCCTCCTGTGCACCCTCTTCATCAGTCTGCGNTCCTCAGACCACCGG
3039067	125	-CTTCCTCCTGTGCACCCTCTTCATCAGTCTGCGCTCCTCAGACCACCGG
2376768	117	-CTTCCTCCTGTGCACCCTCTTCATCAGTCTGCGCTCCTCAGACCACCGG
1682539	107	-CTTCCTCCTGTGCACCCTCTTCATCAGTCTGCGCTCCTCAGACCACCGG
1804118	104	-CTTCCTCCTGTGCACCCTCTTCATCAGTCTGCGCTCCTCAGACCACCGG
2193380	89	-CTTCCTCCTGTGCACCCTCTTCATCAGTCTGCGCTCCTCAGACCACCGG
3210913	90	-CTTCCTCCTGTGCACCCTCTTCATCAGTCTGCGCTCCTCAGACCACCGG
2321125	76	-CTTCCTCCTGTGCACCCTCTTCATCAGTCTGCGCTCCTCAGACCACCGG
2743359	69	-CTTCCTCCTGTGCACCCTCTTCATCAGTCTGCGCTCCTCAGACCACCGG
1755244	69	-CTTCCTCCTGTGCACCCTCTTCATCAGTCTGCGCTCCTCAGACCACCGG
1752950	69	-CTTCCTCCTGTGCACCCTCTTCATCAGTCTGCGCTCCTCAGACCACCGG
3255760	66	-CTTCCTCCTGTGCACCCTCTTCATCAGTCTGCGCTCCTCAGACCACCGG
1655567	57	-CTTCCTCCTGTGCACCCTCTTCATCA
AA477170.RC	57	-CTTCCTCCTGTGCACCCTCTTCATCAGTCTGCGCTCCTCAGACCACCGG
H02520	45	-CTTCCTCCTGTGCACCCTCTTCATCAGTCTGCGCTCCTCAGACCACCGG

FIGURE 4N

```
2623580          34  -CTTCCTCCTGTGCACCCTCTTCATCAGTCTGCGCTCCTCAGACCACCGG
732580           27  -NTTCCTNCTGTGCACCCTCTTCATCAGTCTGCGNTCCTNAGACCACCGG
1579688          30  -CTTCCTCCTGTGCACCCTCTTCATCAGTCTGCGCTCCTCAGACCACCGG
1579888          30  -CTTCCTCCTGTGCACCCTCTTCATCAGTCTGCGCTCCTCAGACCACCGG
794186           15  -CTTCCTCCTGTGCANCCTCTTCATCAGTCTGCGCTCCTCAGACCANCGG
1733505          11  -CTTCCTCCTGTGCACCCTCTTCATCAGTCTGCGCTCCTCAGACCACCGG
1450343           1          CTGAGAGCCCAGGGGCACAGTCTGCGCTCCTCAGACCACCGG
724923            1                           .TGCAGTCTGCGNTCCTNAGACCACCGG
                      .++.++.++..+....+........+++++.++.+.++.+.........
<consen01>     1036   CTTCCTCCTGTGCACCCTCTTCATCAGTCTGCGCTCCTCAGACCACCGG
DNA35930.init   678  GGACGGGG-C--TG-CTGGAGAGAGCGGGGAACTCCA
W49844          464  GGACGGGG-C--TG-CTGGAGAGAGCGGGGAACTCCANCACAGTGGG
274786          248  CAGGTGAA-C--AG-ACTGTTNCAGACCNAGGAGTGNCCACNTATGCTAN
AA130551        237  CAAGNATA-CCAAG-CCTGATGCAGACCGAGGAGTGCCCACCTATGCTAG
AA481928.RC     241  CAGGTGAA-C--AG-CCTGATGCAGACCGAGGAGTGCCCACCTATG-TAG
023897          230  CAGGGAAATC--AGACNTTATGTAGATCGAGGTGTGTCNANNTTAT
1453117         229  CAGGTGAA-C--AG-CCTGATGCAG
1731292         212  CAGGTGAA-C--AG-CCTGA
2417559         207  CAGGTGAA-C--AG-CCTGATGCAGACCGAGGAGTGC
AA410947.RC     203  CAGGTGAA-C--AG-CCTGATGCAGACCGAGGAGTGCCCACCTATGCTAG
1754776         200  CAGGTGAA-C--AG-CCTGATGCAGACCGAGGAGTGCCCACCT
2279337         189  CAGGTGAA-C--AG-CCTGATGCAGACCGAGGAGTGCCCACCTATGCTAG
941498          189  CAGGTGAA-C--AG-CCTGATGCAGACCGAGGAGTGCCCACCTATGCTAG
755392          184  CAGGTGAA-C--AG-CCTGATGCAGACCGAGGAGTGCCCACCTATGCTAG
548789          178  CAGGTGAA-C--AG-CCTGATGCAGACCGAGGAGTGCCCACCTATGCTAG
3039067         174  CAGGTGAA-C--AG-CCTGATGCAGACCGAGGAGTGCCCACCTATGCTAG
2376768         166  CAGGTGAA-C--AG-CCTGATGCAGACCGAGGAGTGCCCACCTATGC
1682539         156  CAGGTGAA-C--AG-CCTGATGCAGACCGAGGAGTGCCCACCTATGCTAG
1804118         153  CAGGTGAA-C--AG-CCTGATGCAGACCGAGGAGTGCCCACCTATGCTAG
2193380         138  CAGGTGAA-C--AG-CCTGATGCAGACCGAGGAGTGCCCACCTATGCTAG
3210913         139  CAGGTGAA-C--AG-CCTGATGCAGACCGAGGAGTGCCCACCTATGCTAG
2321125         125  CAGGTGAA-C--AG-CCTGATGCAGACCGAGGAGTGCCCACCTATGCTAG
2743359         118  CAGGTGAA-C--AG-CCTGATGCAGACCGAGGAGTGCCCACCTATGCTAG
1755244         118  CAGGTGAA-C--AG-CCTGATGCAGACCGAGGAGTGCCCACCTATGCTAG
1752950         118  CAGGTGAA-C--AG-CCTGATGCAGACCGAGGAGTGCCCACCTATGCTAG
```

FIGURE 40

```
3255760          115 CAGGTGAA-C--AG-CCTGATGCAGACCGAGGAGTGCCCACCTATGCTAG
AA477170.RC      106 CAGGTGAA-C--AG-CCTGATGCAGACCGAGGAGTGCCCACCTATG-TAG
H02520            94 CAG-TANA-C--AG-CCTGATGCAGACCGAGGAGTGCCCACCTATGCTAG
2623580           83 CAGGTGAA-C--AG-CCTGATGCAGACCGAGGAGTGCCCACCTATGCTAG
732580            76 CAGGTGAA-C--AG-CCTGATGCAGACCGAGGAGTGCCCACCTATGCTAG
1579688           79 CAGGTGAA-C--AG-CCTGATGCAGACCGAGGAGTGCCCACCTATGCTAG
1579888           79 CAGGTGAA-C--AG-CCTGATGCAGACCGAGGAGTGCCCACCTATGCTAG
794186            64 CAGGTGAA-C--AG-CCTGATGCAGACCGAGGAGTGCCCACCTATGCTAG
1733505           60 CAGGTGAA-C--AG-CCTGATGCAGACCGAGGAGTGCCCACCTATGCTAG
1450343           43 CAGGTGAA-C--AG-CCTGATGCAGACCGAGGAGTGCCCACCTATGCTAG
724923            28 CAGGTGAA-C--AG-CCTGATGCAGACCGAGGAGTGCCCANCTATGCTAG
                     ........ +  .+ ........++.............+.......++.
<consen01>      1085 CAGGTGAA C  AG CCTGATGCAGACCGAGGAGTGCCCACCTATGCTAG

274786           294 ACGGCACAC
AA130551         285 ACGCCACACAGCAGCAG---CAGCAGCAGGTGGCAAGCCTGTGAGGGCCG
AA481928.RC      286 ACGCCACACAGCAGCAG---CAGCAG---GTGGCA-GCCTGTGAGGGCCG
AA410947.RC      249 ACGCCACACAGCAGCAG---CAGCAG---GTGGCA-GCCTGTGAGGGCCG
2279337          235 ACGCCACA
941498           235 ACGCCACA
755392           230 ACGNCAC
548789           224 ACGGCACACAGCAGCAG---CAG
3039067          220 ACGCCACA
1682539          202 ACGCCACA
1804118          199 ACGCCACA
2193380          184 ACGCCACA
3210913          185 ACGCCACACAGCAGCAG---CAGCAG---GTGGCA-GCCTGTGAGGGCCG
2321125          171 ACGCCACA
2743359          164 ACGCCACA
1755244          164 ACGCCACANNNNNNNNN---NNNNNN---GTGGCA-GCCTGTGAGGGCCG
1752950          164 ACGCCACANNNNNNNNN---NNNNNN---GTGGCA-GCCTGTGAGGGCCG
3255760          161 ACGCCACANNNNNNNNN---NNNNNNNNNGTGGCA-GCCTGTGAGGGCCG
AA477170.RC      151 ACGCCACACAGCAGCAG---CAGCAG---GTGGCA-GCCTGTGAGGGCCG
H02520           139 ACGCCACACAGCAGCAG---CAGCAGCAGGTGGCA-GCCTGTGAGGGCCG
2623580          129 ACGCCACANNNNNNNNN---NNNNNNNNNGTGGCA-GCCTGTGAGGGCCG
```

FIGURE 4P

```
732580          122 ACGNCACANNNNNNNNN---NNNNNNNNGTGGCA-GCCTGTGAGGGCCG
1579688         125 ACGCCACA
1579888         125 ACGCCACA
794186          110 ANGGCACANNNNNNNNN---NNNNNNNNGTGGCA-GCCTGTGAGGGCCG
1733505         106 ACGCCACANNNNNNNNN---NNNNNNNNGTGGCA-GCCTGTGAGGGCCG
1450343          89 ACGCCACANNNNNNNNN---NNNNNNNNGTGGCA-GCCTGTGAGGGCCG
724923           74 ACGNCACANAGNAGCAGNAGCAGCAG---GTGGCA-GCCTGTGAGGGCCG
1310973           1                             GTGGCA-GCCTGTGAGGGCCG
836342            1                             GTGGCA-GCCTGTGAGGGCCG
836297            1                             GTGGCA-GCCTGTGAGGGCCG
1802601           1                             GTGGCA-GCCTGTGAGGGCCG
1529466           1                                 CA-GCCTGTGAGGGCCG
                    +.+.++++ .. .....   ...     ++++++ ++++++++++++++
<consen01>     1131 ACGCCACACAGCAGCAG   CAGCAG   GTGGCA GCCTGTGAGGGCCG

AA130551        332 GGCCTTTGACAACGAGCAGGACGGCGTCACCTACAGCTAC-TCCTTCTTC
AA481928.RC     329 GGCCTTTGACAACGAGCAGGACGGCGTCACCTACAGCTAC-TCCTTCTTC
AA410947.RC     292 GGCCTTTGACAACGAGCAGGACGGCGTCACCTACAGGGAGGTGGCCCGGG
3210913         228 GGCCTTTGACAACGAGCAGGACGGCGTCACCTACAGCTAC-TCCTTCTTC
1755244         207 GGCCTTTGACAACGAGCAGGACGGCGTCACCTACAGCTAC
1752950         207 GGCCTTTGACAACGAGCAGGACGGCGTCACCTACAG-TAC-TCCT
3255760         207 GGCCTTTGACAACGAGCAGGACGGCGTCACCTACAGCT
AA477170.RC     194 GGCCTTTGACAACGAGCAGGACGGCGTCACCTACAGCTAC-TCCTTCTTC
H02520          185 GGCCTTTGACAACGAGCAGGACGGCGTCACCTACAGCTAC-TCCTTCTTC
2623580         175 GGCCTTTGACAACGAGCAGGACGGCGTCACCTACAGCTAC-TCCTTCTTC
732580          168 GGCCTTTGACAACGAGCAGGACGGCGTCANCTACAGNTAC-TNCTTNTTC
794186          156 GGCCTTTGACAANGAGCAGGACGGNGTCANCTACAGCTAC-TNCTTCTTC
1733505         152 GGCCTTTGACAACGAGCAGGACGGCGTCACCTACAGCTAC-TCCTTCTTC
1450343         135 GGCCTTTGACAACGAGCAGGACGGCGTCACCTACAGCTAC-TCCTTCTTC
724923          120 GGCCTTTGACAACGAGCAGGACGGCGTCAACTACAGNTAC-TTCTTGTTN
1310973          21 GGCCTTTGACAACGAGCAGGACGGCGTCACCTACAGCTAC-TCCTTCTTC
836342           21 GGCCTTTGACAACGAGCAGGACGGCGTCACCTACAGCTAC-TCCTTCTTC
836297           21 GGCCTTTGACAACGAGCAGGACGGCGTCACCTACAGCTAC-TCCTTCTTC
1802601          21 GGCCTTTGACAACGAGCAGGACGGCGTCACCTACAGCTAC-TCCTTCTTC
1529466          17 GGCCTTTGACAACGAGCAGGACGGCGTCACCTACAGCTAC-TCCTTCTTC
```

FIGURE 4Q

```
238109              1                    GGACGGCGTCACCTACAGCTAC-TCCTTCTTC
2528337             1                                                   TC
                     ++++++++++++.++++++++++.++++.++++++..+. +........
<consen01>       1174 GGCCTTTGACAACGAGCAGGACGGCGTCACCTACAGCTAC TCCTTCTTC

AA130551         381 CACTTCTGCCTGG-TGCTGGCCTCACTG-CACGTCATGATGACGCTCACC
AA481928.RC      378 CACTTCTGCCTGG-TGCTGGCCTCACTG-CACGTCATGATGACGCTCACC
AA410947.RC      342 GGCGTCTGCCTGG-TGCTGGCCTCACTG-CACGTCATGATGACGCTCACC
3210913          277 CA-TTCTGCCTGG-TGCTGGCCTCACTG-CAGT
AA477170.RC      243 CACTTCTGCCTGG-TGCTGGCCTCACTG-CACGTCATGATGACGCTCACC
H02520           234 CACTTCTGCCTGGGTGCTGGCCTCACTGGCACGTCATGATGACGCTCACC
2623580          224 CACTTCTGCCTGG-TGCTGGCCTCACTG-CACGTCATG
732580           217 CACTTNTGCCTGG-TGCT
794186           205 CACTTCTGCCTGG-TGCTGGNCTNACTG-CACGTNA
1733505          201 CACTTCTGCCTGG-TG
1450343          184 CACTTCTGCCTGG-TGCTGGCCTCACTG-CACGTCATGATGACGCTCACC
724923           169 CANTTNTGNCTGG-TGCTGGGCTNANTG-CACGTTATGATGACGGTNAAC
1310973           70 CACTTCTGCCTGG-TGCTGGCCTCACTG-CACGTCATGATGACGCTCACC
836342            70 CACTTCTGCCTGG-TGCTGGCCTCACTG-CACGTCATGATGACGCTCACC
836297            70 CACTTCTGCCTGG-TGCTGGCCTCACTG-CACGTCATGATGACGNTCANC
1802601           70 CACTTCTGCNTGG-NGCTGGCCTCACTG-CACGNCATGATGACGCTCACC
1529466           66 CACTTCTGCCTGG-TGCTGGCCTCACTG-CACGTCATGATGACGCTCACC
238109            32 CACTTCTGCCTGG-TGCTGGCCTCACTG-CACGTNATGATGACGCTCACC
2528337            3 CACTTCTGCCTGG-TGCTGGCCTCACTG-CACGTCATGATGACGCTCACC
1228220            1   CTTCTGCCTGG-TGCTGGCCTCACTG-CACGTCATGATGACGCTCACC
900944             1                          CACTG-CACGTCATGATGACGCTCACC
900010             1                          CACTG-CACGTCATGATGACGCTCACC
877907             1                                        GACGCTCACC
2252522            1                                              CACC
2272202            1                                              CACC
                     ....+.++..+++ .+++++.++.+.++ ++....+++++++++.+.+.+
<consen01>       1223 CACTTCTGCCTGG TGCTGGCCTCACTG CACGTCATGATGACGCTCACC
```

FIGURE 4R

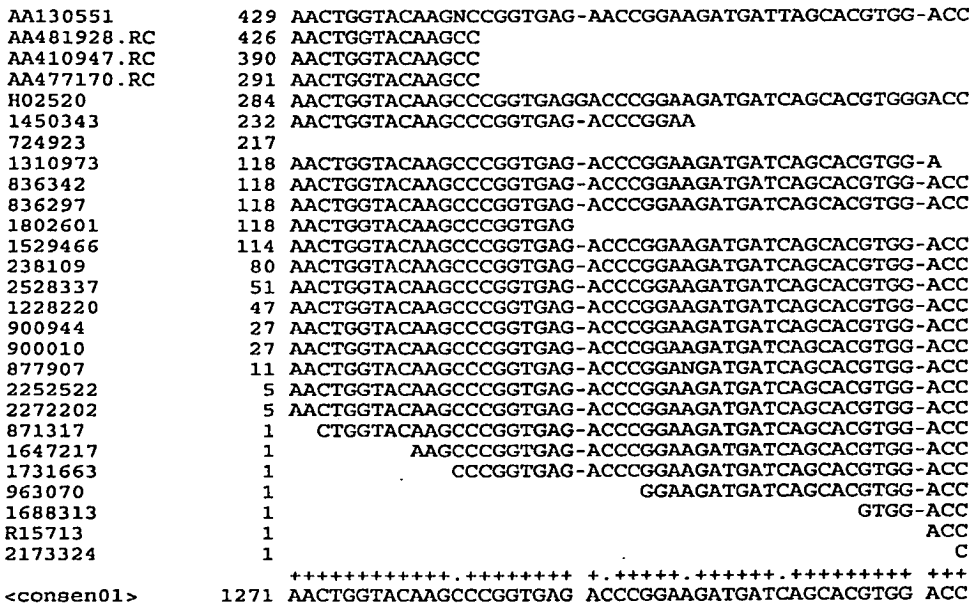

FIGURE 4S

```
AA130551    477 GCCGTGT--GGGTGAAGATTTTGT-GCCAGCT-GGGCAAG-GGCTTGCTC
H02520      334 GCCGTGTNGGGGTGAAGATNT-GTTGCCAGCTNGGGCA-G-GGGTTGTTC
836342      166 NCCGTGT--GGGTGAAGATCT-GT-GCCAGCT-GGGCA-G-GGCT-GCTC
836297      166 GCCGTGT--GGGTGAAGATCT-GT-GCCAGCT-GGGCA-G-GGCT-GCTC
1529466     162 GCCGTGT--GGGTGAAGATCT-GT-GCCAGCT-GGGCA-G-GGCT-GCTC
238109      128 GCCGTGT--GGGTGAAGATCT-NT-GCCAGCT-GGGCA-G-GGCT-GCTC
2528337      99 GCCGTGT--GGGTGAAGATCT-GT-GCCAACT-GGGCA-G-G-CT-GCTC
1228220      95 GCCGTGT--GGGTGAAGATCT-GT-GCCAGCT-GGGCA-G-GGCT-GCTC
900944       75 GCCGTGT--GGGTGAAGATCT-GT-GCCAGCT-GGGCA-G-GGCT-GCTN
900010       75 GCCGTGT--GGGTGAAGATCT-GT-GCCAGCT-GGGCA-G-GGCT-GCTC
877907       59 GCCGTGT--GGGTGAAGATCT-GT-GCCAGCT-GGGCA-G-GGCT-GCTG
2252522      53 GCCGTGT--GGGTGAAGATCT-GT-GCCAGCT-GGGCA-G-GGCT-GCTT
2272202      53 GCCGTGT--GGGTGAAGATCT-GT-GCCAGCT-GGGCA-G-GGCT-GCTC
871317       47 GCCGTGT--GGGTGAAGATCT-GT-GCCANCT-GGGCA-G-GGCT-GCTC
1647217      40 GCCGTGT--GGGTGAAGATCT-GT-GCCAGCT-GGGCA-G-GGCT-GCTC
1731663      37 GCCGTGT--GGGTGAAGATCT-GT-GCCAGCT-GGGCA-G-GGCT-GCTC
963070       24 GCCGTGT--GGGTGAAGATCT-GT-GCCAGCT-GGGCA-G-GGCT-GCTC
1688313       8 GCCGTGT--GGGTGAAGATCT-GT-GCCAGCN-GGGCA-G-GGCT-GCTC
R15713        4 GCCGTGT--GGGTGAAGATCT-GT-GCCAGCT-GGN-A-GAGGCT-GCTC
2173324       2 GCCGTGT--GGGTGAAGATCT-GT-GCCAGCT-GGGCA-G-GGCT-GCTC
H00901        1                   TCT-GT-GCCAGCT-GGGCA-G-GGCT-GCTC
R67665        1                   TCT-GT-GCCAGCT-GGG-A-G-GGCT-GCTC
2840414       1                       GT-GCCAGCT-GGGCA-G-GGCT-GCTC
1315226       1                               GCT-GGGCA-G-GGCT-GCTC
1454892       1                                           GGCT-GCTC
R73342        1                                             CT-GCTC
R73357        1                                             CT-GCTC
1981311       1                                                 CTC
                .++++++  ++++++++++.+ .+ ++++.+. ++.++ + ++.+ +.+.
&lt;consen01&gt;   1319 GCCGTGT  GGGTGAAGATCT GT GCCAGCT GGGCA G GGCT GCTC
```

FIGURE 4T

```
AA130551          522 C
H02520            381 CT-NTTACCTGTTGGACCTGGGGTAAGGCCCCAT
836342            208 CT-CT-ACCTGT-GGACCCTGGTAGCCCCACTTCTTCTGCGCAACCGCG-
836297            208 CT-CT-ACCTGT-GGACCCTGGTAGCCCCACTNCTNCTGCGCAACCGNG-
1529466           204 CT-C
238109            170 CT-CT-ACCTGT-GGACCCTGGTAGCCCCACTCCTCCTGCGCAACCGCG
2528337           140 CT-CT-AACTGT-GGAACCTGGTAGCCCAATTCCTCCTGCGCAACCG-G-
1228220           137 CT-CT-ACCTGT-GGACCCTGGTAGCCCCACTCCTCCTGCGCAACCGCG-
900944            117 CT-CT-ACCTGT-GGACCCTGGTAGCCCCACTCCTNCTGCGCAACCGNG-
900010            117 CT-CT-ACCTGT-GGACCCTGGTAGCCCCACTCCTCCTGCGCAACCGCG-
877907            101 CT-CT-ACCTGT-GGACCCTGGTAGCCCCANTNCTGCTGCGCAACCGCGG
2252522            95 CTTCT-ACCTGT-GGACCCTGGTAGCCCCANTCCTCCTGCGCAACCGCG-
2272202            95 CT-CT-ACCTGT-GGACCCTGGTAGCCCCACTCCTCCTGCGCAACCGCG-
871317             89 CT-CT-ACCTGT-GGACCCTGGTAGCCCCACTTCTNCTGCGCAACCGCG-
1647217            82 CT-CT-ACCTGT-GGACCCTGGTAGCCCCACTCCTCCTGCGCAACCGCG-
1731663            79 CT-CT-ACCTGT-GGACCCTGGTAGCCCCACTCCTCCTGCGCAACCGCG-
963070             66 CT-CT-ACCTGT-GGACCCTGGTAGCCCCACTNCTCCTGCGCAACCGCG-
1688313            50 CT-CT-ACCTGT-GGACCCTGGTAGCCCCACTCCTCCTGCGCAACCGCG-
R15713             46 CT-CT-ACCTGT-GGACCCTGGTAGCCCCACTCCTCCTGCGCAACCG-G-
2173324            44 CT-CT-ACCTGT-GGACCCTGGNAGCCCCACTCCTCCTGCGCAACCGCG-
H00901             27 CT-CT-ACCTGT-GGACCCTGGTAGCCCCACTCCTCCTGCGCAACCGCG-
R67665             26 CT-CT-ACCTGT-GGACCCTGGTAGCCCCACTCCTCCTGCGCAACCG-G-
2840414            24 CT-CT-ACCTGT-NGACCCTGGTAGCCCCACTCCTCCTGCGCAACCGCN-
1315226            18 CT-CT-ACCTGT-GGACCCTGGTAGCCCCACTCCTCCTGCGCAACCGCG-
1454892             9 CT-CT-ACCTGT-GGACCCTGGTAGNCCCACTCCTCCTGCGCAACCGCG-
R73342              7 CT-CT-ACCTGT-GGACCCTGGTAGCCCCACTCCTCCTGCGCAACCGCG-
R73357              7 CT-CT-ACCTGT-GGACCCTGGTAGCCCCACTCCTCCTGCGCAACCGGA-
1981311             4 CT-CT-ACCTGT-GGACCCTGGTAGCCCCACTCCTCCTGCGCAACCGCG-
1518815             1                                          TGCGCAACCGCG-
AA211365.RC         1                                                CCGGCG-
1734407             1                                                 CCGCG-
2620380             1                                                  CGCG-
2131632             1                                                   GCG-
                      ++ .+ +.++++ .++.+..++............+.+++++++.+.+..
<consen01>        1361 CT CT ACCTGT GGACCCTGGTAGCCCCACTCCTCCTGCGCAACCGCG
```

FIGURE 4U

```
836342         254 A-C-TT-CAGCTGAGGCAGC-T-NACAGC-TGC-CATCTGGTGCTTCTGC
836297         254 A-C-TT-CAGCTGAGGCAGCCT-CACAGCCTGC-CATCTGGTGCCTTCTG
2528337        185 A---TT-CA-CTGAGGCAGC-T-CACAGCCTGC-NATCTGGTGC-TCCTG
1228220        183 A-C-TT-CAGCTGAGGCAGCCT-NACAGCCTGC-CATCTGGTGCCTGCTG
900944         163 A-C-TT-CAGCTGAGGCAGCCT-CACAGCCTGC-CATCTGGTGCCTACTG
900010         163 A-C-TT-CAGCTGAGGCAGCCT-CACAGCCTGC-CATCTGGTGCCTCCTG
877907         148 A-C-TT-CAGCTGAGGCAGCCTTNACAGCCTGCNCATCTGGTGCCTNCTG
2252522        142 A-C-TT-CAGCTGAGGCAGCCT-CACAGCCTGC-CATCTGGTGCCTCCTG
2272202        141 A-C-TT-CAGCTGAGGCAGCCT-CACAGCCTGC-CATCTGGNGCCTCCTG
871317         135 A-C-TT-NAGCTGAGGCAGCCT-NACAGCCTGC-CATCTGGTGCCTNCTG
1647217        128 A-C-TT-CAGCTGAGGCAGCCT-CACAGCCTGC-CATCTGGTGCCTCCTG
1731663        125 A-C-TT-CAGCTGAGGCAGCCT-CACAGCCTGC-CATCTGGTGCCTCCTG
963070         112 A-C-TT-CAGCTGAGGCAGCCT-CACAGCCTGC-CATCTGGTGCCTNCTG
1688313         96 A-C-TT-CAGCTGAGGCAGCCT-CACAGCCTGC-CATCTGGTGCCTCCTG
R15713          91 A-C-TTTCAGCTGAGGCAGCCT-CACAGCCTGC-CATCTGGTGCCTCCTG
2173324         90 A-C-TT-CAGCTGANGCAGCCT-CACAGCCTGC-CATCTGGTGCCTCCTG
H00901          73 A-C-TT-CAGCTGAGGCAGCCT-CACAGCCTGC-CATCTGGTGCCTCCTG
R67665          71 A-CCTT-CAGCTGAGGCAGCCT-CACAGCCTGC-CATCTGGTGCCTCCTG
2840414         70 A-C-TT-CAGCTGANGCAGCCT-CACAGCCTGC-CATCTGGNGCCTCCTG
1315226         64 A-C-TT-CAGCTGAGGCAGCCT-CACAGCCTGC-CATCTGGTGCCTCCTG
1454892         55 A-C-TT-CAGCTGAGGCAGCCT-CACAGCCTGC-CATCTGGTGCCTCCTG
R73342          53 ANC-TT-CAGCTGAGGCAGCCT-CACAGCCTGC-CATCTGGTGCCTCCTG
R73357          53 N-C-TT-CAGCTGAGGCAGCCT-CACAGCCTGC-CATCTGGTGCCTCCTG
1981311         50 A-C-TT-CAGCNGAGGCAGCCT-CACAGCCTGC-CATCTGGTGCCTCCTG
1518815         13 A-C-TT-CAGCTGAGGCAGCCT-CACAGCCTGC-CATCTGGTGCCTCCTG
AA211365.RC      7 A-C-TTTCAGCTGAG-CAGCTT-CACAGCCTGC-AATCTGGTGCCTCCTG
1734407          6 A-C-TT-CAGCTGAGGCAGCCT-CACAGCCTGC-CATCTGGTGCCTCCTG

2620380          5 A-C-TT-CAGCTGAGGCAGCCT-CACAGCCTGC-CATCTGGTGCCTCCTG
2131632          4 A-C-TT-CAGCTGANGCAGCCT-CACAGCCTGC-NATCTGGTGCCTCCTG
AA130587.RC      1                       TT-CACAGNCTGC-CATCTGGTGCCTCCTG
R72869.RC        1                                                 CTTG
237971           1                                               CTCCTG
                   . + ++ .+++.++.+++++.+ .++++.++++ .++++++.++.+....
<consen01>    1407 A C TT CAGCTGAGGCAGCCT CACAGCCTGC CATCTGGTGCCTCCTG
```

08622480" 05428005

FIGURE 4V

836342	297	C
836297	299	CCA-NCTGGT-GCCT-NTNGGNT-CGGT
2528337	226	CCA-C-TGGTTGCTT-NTNGNTT-CGGTGA-AAGCAAN-CN-GGNANCNN
1228220	228	CCA-CCTGGT-GCCT-NTCGGCT-NAA
900944	208	NCA-CCTGGT-GCCT-CTNGGCT-CGGTGA-CAGCCAA-CC-TNANCCNN
900010	208	CCA-CCTGGT-GCCT-CTCGGCT-CGGTGA-CAGCCAA-CC-TNCCCCCT
877907	195	CCA-CCTGGT-GCCT-GTNGGCTTCGGTGAACAGCCAAACCNTNCCGGCT
2252522	187	CCA-CCTGGT-GCCT-CTCGGCT-CGGTGA-CAGCCAA-CN-GNNCCNCN
2272202	186	CCA-CCTGGN-GCCT-CTNGGCT-CGGTGA-CAGCNAA-CC-TNNNCCCT
871317	180	CCA-CCTGGT-GCCT-CTNGGCT-CGGTGA-CAGCCAA-NC-TGNCCCCT
1647217	173	CCA-CCTGGT-GCCT-CTCGGCT-CGGTGA-CAGCCGA-NN-NNANNNNN
1731663	170	CNA-CCTGGT-GCCT-CTCGGCT-CGGTGA-CAGCCAA-CN-GNNNC
963070	157	CCA-CCTGGT-GCCT-GTCGGCT-CGGTGA-CAGCCAA-C--TGNNCCNT
1688313	141	CCA-CCTGGT-GCCT-CTCGGCT-CGGTGA-CAGCCAA-CC-NTNNACCN
R15713	137	CCA-CCTGGT-GCCT-CTCGGCT-CAGTGA-CAGCCAA-CC-TGCCCCCT
2173324	135	CCA-CCTGGN-GCCT-CTNGGCT-CGGTGA-CAGCNAA-CC-AGNGCCCT
H00901	118	CCA-CCTGGT-GCCT-CTCGGCT-CGGTGA-CAGCCAA-CC-TGCCCCCT
R67665	117	CCA-CCTGGT-GCCT-CTCGGCT-CGGTGA-CAGCCAA-CC-TGCCCCCT
2840414	115	C
1315226	109	CCA-CCTGGT-GCCT-CTCGGCT-CGGTGA-CAGCCAA-CC-NNCNCCCT
1454892	100	CCA-CCTGGT-GCCT-CTCGGNT-CGGNGA-CAGCCAA-CC-NNCCCCCT
R73342	99	CCA-CCTGGT-GCCT-CTCGGCT-CGGTGA-CAGCCAA-CC-TGCCCCCT
R73357	98	CCA-CCTGGT-GCCT-CTCGGCT-CGGTGA-CAGCCAA-CC-TGCCCCCT
1981311	95	CCA-CCTGGT-GCCT-CTCGGCT-CGGTGA-CAGCCAA-CC-NG
1518815	58	CCA-CCTGGT-GCCT-CTCGGCT-CGGTGA-CAGCCAA-CC-TNNNCCCT
AA211365.RC	52	CCA-CCTTGG-GTNTTCTCGGCT-CGGTGA-CAGCCAA-CC-TGCCCCCT
1734407	51	CCA-CCTGGT-GCCT-CTCGGCT-CGGTGA-CAGCCAA-CC-NNNNNCCT
2620380	50	CCA-CCTGGT-GCCT-CTCGGCT-CAGTGA-CAGCCAA-CN-NAGNCNNN
2131632	49	CNA-CCTGGT-GCCT-CTNGGCT-CGGTGA-CAGCNAA-CC-NNANCCN-
AA130587.RC	29	CCAACNTGG--GTGT-CTCGGCT-CGGTGA-CAGCCAA-CC-TGCCCCCT
R72869.RC	5	GNT-TTTCTG-GGCT-GGGTTNA-AAAGCN-AAATTGC-NC-CTTNCCCC
237971	7	CCA-CCTGGT-GCCT-CTGGGCT-CGGTGA-CAGCCAA-CC-TGCCCCCT
238502	1	CCA-CCTGGT-GCCT-CTNGGCT-CGGTGA-CAGCCAA-CC-TGCCCCCT
R72008.RC	1	ATTG-CN-CCTTCNCC
AA480065.RC	1	CCCCCT

FIGURE 4W

```
                      ...  ..+... +..+ ....... ...... .+.... .. .  .....
<consen01>       1452 CCA CCTGGT GCCT CTCGGCT CGGTGA CAGCCAA CC TGCCCCCT

2528337           269 -ANCAANAN-AAN-TAANCNAG-NNNNAN-ANGNA-NNNNTN--NCN-NA
900944            251 -NCCCACAC-CAA-TCAGCCAG-GCTGAG-CCCCC-ACCCCT--GCC-CC
900010            251 -CCCCACAC-CAA-TCAGCCAG-GCTGAG-CCCCC-ACCCNT--GCC-CC
877907            242 TCCNNACAAGCAANTNAGCCAGNGCTGAG-CCCCC-ACCNCTT-GCC-NG
2252522           230 -GGNCANAN-CAA-TCAGCCAG-GCT
2272202           229 -CCCCANAC-CAA-TCAGCCAG-GCTGAN-CNCCC-ACCCCT--GCN-CC
871317            223 -TCCCACAA-CAA-TNAGCCAG-GGT
1647217           216 -CCNA
963070            199 -TCCCANAA-CAA-TTAGNCAG-GTTGGN-CCCCA-ACACCT--GCC-C-
1688313           184 -NNCCANAC-CAA-TNAGCCAG-GCTGAN-ANNCC-ANCCCT--NNCNCC
R15713            180 -CCCCACAC-CAA-TCAGCCAG-GCTGAG-CCCCC-ACCCCT--GCC-CC
2173324           178
H00901            161 -CCCCACAC-CAA-TCAGCCAG-GCTGAG-CCCCC-ACCCCT--GCC-CC
R67665            160 -CCCCACAC-CAA-TCAGCCAG-GCTGAG-CCCCC-ACCCCT--GCC-CC
1315226           152 -CCCCACAC-CAA-TCNGCCAG-GCTGAN-CNCCC-ACCCCT--GCC-CC
1454892           143 -CNCCANAC-CAN-TCAGCCAGCGCTGAN-CCCCC-ACCCCT--GCN-NC
R73342            142 -CCCCACAC-CAA-TCAGCCAG-GCTGAG-CCCCC-ACCCCT--GCC-CC
R73357            141 -CCCCACAC-CAA-TCAGCCAG-GCTGAG-CCCCC-ACCCCT--GCC-CC
1518815           101 -CCCCACAC-CAA-TCAGCCAG-GCTGAN-CNCCC-ACCCCT--GCC-CC
AA211365.RC        96 -CCCCACAC-CAA-TCAGCCAG--CTGAG-CCCCC-ACCCCT--GCC-CC
1734407            94 -CCCCANAC-CAA-TCAGCCAG-GCTGAN-CNCCC-ACCCCT--GCN-CC
2620380            93 -NCNCANAC-CAN-TNAGCCAG-GCTGANACNCCC-ANNCCT--GNN-CC
2131632            91 --CCNNNAC--AA-TNAGCCAG-GCTGAN-NGCCA-CCCTGN--CCA-GT
AA130587.RC        72 -CCCCACAC-CAA-TCAGCCAG--CTGAG-CCCCC-ACCCCT--GCC-CC
R72869.RC          48 -AAAAACNA-AAT-TAAGCCAA-GNTGAG-CCCCC-AACCCTTTGCC-CC
237971             50 -CCCCACAC-CAA-TCAGCCAG-GCTGAG-CCCCC-ACCCCT--GCC-CC
238502             44 -CCCCACAC-CAA-TCAGCCAG-GCTGAG-CCCCC-ACCCCT--GCC-CC
R72008.RC          15 -CAAAAACN-AAA-TAAGCCNA-GCTGAG-CCCCCNACCCCT--GCC-CA
AA480065.RC         7 -CCCCACAC-CAA-TCAGCCAG--CTGAG-CCCCC-ACCCCT--GCC-CC
1818182             1        CAC-CAA-TCAGCCAG-GCTGAN-CCCCC-ACCCCT--GCC-CC
2194762             1        CAC-CAA-TCAGCCAG-GCTGAN-CCCCC-ACCCCT--GCC-CC
1818702             1        CAC-CAA-TCAGCCAG-GCTGAN-CCCCC-ACCCCT--GCC-CC
```

FIGURE 4X

```
AA010655          1       CAC-CAA-TCAGCCAG-GCTGAN-CCCCC-NCCCCT--GCC-CC
AA010656.RC       1                                  CNC-ACCCCT--GCC-CC
R16077.RC         1                                            G--NCC-CA
R93337.RC         1                                           CT--GCC-CC
2053751           1                                               GCN-CC
H00902.RC         1                                                   CC
                         ........ .+. +....... ...... ..... ......  ... ..
<consen01>     1495      CCCCACAC CAA TCAGCCAG GCTGAG CCCCC ACCCCT  GCC CC

2528337         310 A-NT
900944          292 A-GT
900010          292 A-GCT-C-C-AGG-A-CCTGCCCCT-GA-G-CCGGG
877907          288 A-GNT-T-CGAGG-AACCTGNCCNTTGA-G-CCGGGCCTT-G-TA-GTCG
2272202         270 A-GCT-C-C-AGG-A-CCTGCCANT
963070          239 A-GTTTCTC-AGG-A-NCTGCCCCT-GA-G-CCGGGNCTT-N-TA-GTTG
1688313         226 A-GCT-C-C-AGG-A-CCTGC
R15713          221 A-GCT-C-C-AGG-A-CCTGCCCCT-GA-G-CCGGGCCTT-C-TA-GTCG
H00901          202 A-GCT-C-C-AGGGA-CCTGCCCCT-GA-G-CCGGGCCTT-C-TA-GTCG
R67665          201 A-GCT-C-C-AGG-A-CCTGCCCCT-GA-G-CCGGGCCTT-C-TA-GTCG
1315226         193 A-GCT-C-C-AGG-A-CCTGCCCCT-GA-G-CCGGGCCTT-C-TA-GTCG
1454892         185 A-GCT-N-C-AGG-A-CCTGCCCNT-GA-G-CCGGGCCTT-C-TA-GTNG
R73342          183 A-GCT-C-C-AGG-A-CCTGCCCCT-GA-G-CCGGGCCTT-C-TA-GTCG
R73357          182 A-GCT-C-C-AGG-A-CCTGCCCCT-GA-G-CCGGGCCTT-C-TA-GTCG
1518815         142 A-GCT-C-C-AGG-A-CCTGCCCCT-GA-G-CCGGGCCTT-C-TA-GTCG
AA211365.RC     136 A-GCT-C-C-AGG-A-CCTGCCCCT-GA-G-CCGGGCCTT-C-TA-GTCG
1734407         135 A-GCT-C-C-AGG-A-CCTGCCCCT-GA-G-CCGGGCCTT-C-TA-GTCG
2620380         135 A-GCT-C-C-AGG-A-CCTGCCCCT-GA-G-NTGGGCCTT-C-TA-GTCG
2131632         130 C-AGG-A-N-TGN-C-TGAGCCGGG-CT-T-TAGTGTAGT-N-CT-TANG
AA130587.RC     112 A-GCT-C-C-AGG-A-CCTGCCCCT-GA-G-CCGGGCCTT-C-TA-GTCG
R72869.RC        91 AAGNT-C-CCAGG-A-CCTGCCCNT-GAAG-CCGGGCCTT-C-TAAGTCG
237971           91 A-GCT-C-C-AGG-A-CCTGCCCCT-NA-G-CCGGGCCTT-C-TA-GTCG
238502           85 A-GCT-C-C-AGG-A-CCTGCCCCT-GA-G-CCGGGCCTT-C-TA-GTCG
R72008.RC        57 A-GNT-C-CCAGG-A-CCTGCCCCT-GAAG-CCGGGCCTT-CCTA-GTCG
AA480065.RC      47 A-GCT-C-C-AGG-A-CCTGCCCCT-GA-G-CCGGGCCTT-C-TA-GTCG
1818182          37 A-GCT-C-C-AGG-A-CCTGCCCCT-GA-G-CCGGGCCTT-C-TA-GTCG
```

FIGURE 4Y

```
2194762          37 A-GCT-C-C-AGG-A-CCTGCCCCT-GA-G-CCGGGCCTT-C-TA-GTCG
1818702          37 A-GCT-C-C-AGG-A-CCTGCCCCT-GA-G-CCGGGCCTT-C-TA-GTCG
AA010655         37 A-GCT-C-C-AGG-A-CCTGCCCCT-GA-G-CCGGGCCTT-C-TA-GTCG
AA010656.RC      15 A-GNT-C-C-AGG-A-CCTGCCCCT-GA-G-CCGGGCCTT-C-TA-GTCG
R16077.RC         7 A-GCT-C-CCAGG-A-CCTGNCCCT-GA-N-CCGGGCCTTNC-TAANTCG
R93337.RC         8 AAGCT-C-CCAGG-A-CCTGCCCCT-GAAG-CCGGGCCTT-C-TA-NTCG
2053751           6 A-NCT-C-C-ANG-A-CCTGCCCCT-GA-G-CCGGGCCTT-C-TA-GTCG
H00902.RC         3 A-GCT-C-CCAGG-A-CCTGCCCCT-GA-AACCGGGCCTT-C-TA-ATCG
R72882.RC         1   TTC-C-C-AGN-A-CCTGCCCCT-GA-G-CCGGGCCTT-C-TAAGTCG
AA434486          1       C-C-AGG-A-CCTGCCCCT-GA-G-CCGGGCCTT-C-TA-GTCG
N94336.RC         1                                  GGCCTTTC-TAAGTCG
2758102           1                                CCGGGCCTT-C-TA-GTCG
                    . ... . . ... . ...+.+... .. . ..+.+...+ . .. ...+
<consen01>     1536 A GCT C C AGG A CCTGCCCCT GA G CCGGGCCTT C TA GTCG

877907           329 TA-GTGGCCTT-CA--GGGTTN-CG-AC
963070           279 TA-GTT
R15713           259 TA-GTG-CCTT-CA--GGGTCC-GA-GG-AGCAT-CAGGCT-TCTTGCAG
H00901           241 TA-GTG-CCTT-CA--GGGTCC-GA-GG-AGCAT-CAGGCT-CCT-GCAG
R67665           239 TA-GTG-CCTT-CA--GGGTCC-GA-GG-AGCAT-CAGGCT-NCT-GCAG
1315226          231 TA-GTG-CCTT
1454892          223 TN-GTG-CNTT-NN--GGGTCC-GA
R73342           221 TA-GTG-CCTT-CA--GGGTCC-GA-GG-AGCAT-CAGGCT-NCT-GCAG
R73357           220 TA-GTG-CCTT-CA--GGGTCC-GA-GG-AGCAT-CAGGCT-CCT-GCAG
1518815          180 TA-GTG-CCTT-CA--GGGTCC-GA-GG-AGCAT-CAGGCT-CCT
AA211365.RC      174 TA-GTG-CCTT-CA--GGGTCC-GA-GG-AGCAT-CAG-CT-CCT-GCAG
1734407          173 TA-GTG-CCTT-CA--GGGTCC-GA-GG-AGCAT-CAGGCT-CCT-GCAG
2620380          173 TA-GTG-CCTT-CA--GGGTCC-GA-GG-AGCNT-CAGGCT-CCT-GCAG
2131632          168 GC-CGA-GGAN-AT--AGGTCC-TG-AG-TTCCA-TCCCCG-CAC-ANCA
AA130587.RC      150 TA-GTG-CCTT-CA--GGGTCC-GA-GG-AGCAT-CAAGCT-CCT-GCAG
R72869.RC        133 TA-GTG-CCTT-CA--GGGTCC-GAAGG-AGCAT-CAGGCT-CCT-GCAG
237971           129 TA-GTG-CCTT-CA--GGGTCC-GA-GG-AGCAT-CAGGCT-CCT-GCAG
238502           123 TA-GTG-CCTT-CA--GGGTCC-GA-GG-AGCAT-CAGGCT-CCT-GCAG
R72008.RC         98 TAAGTG-CCTT-CA--GGGTCC-GAAGG-AGCAT-NAAGCT-CCT-GCAG
AA480065.RC       85 TA-GTG-CCTT-CA--GGGTCC-GA-GG-AGCAT-CAGGCT-CCT-GCAG
```

FIGURE 4Z

```
1818182          75 TA-GTG-CCTT-CA--GGGTCC-GA-GG-AGCAT-CAGGCT-CCT-GCAG
2194762          75 TA-GTG-CCTT-CA--GGGTCC-GA-GG-AGCAT-CAGGCT-CCT-GCAG
1818702          75 TA-GTG-CCTT-CA--GGGTCC-GA-GG-AGCAT-CAGGCT-CCT-GCAG
AA010655         75 TA-GTG-CCTT-CA--GGGTCC-GA-GG-AGCAT-CAGGCT-CCT-GCAG
AA010656.RC      53 TA-GTG-CCTT-CA--GGGTCC-GA-GG-AGCAT-CAAGCT-CCT-GCAG
R16077.RC        48 TAAGTG-CCTT-CA--GGGTCCCGAAGG-ANCAT-CAAGCT-CCT-GNAG
R93337.RC        49 TA-GTG-CCTT-NA--GGGTCC-GA-GG-AGCATTCAGGCT-CCT-GCAG
2053751          44 TA-GTG-CCTT-CA--GGGTCC-GA-GG-AGCAT-CAGGCT-CCT-GCAG
H00902.RC        43 TA-GTG-CCTTTCA--GGGTCC-GA-GG-AGCATTCAGGCT-CCT-GCAN
R72882.RC        39 TAAGTG-CCTT-CA--GGGTCC-GAAGG-AGCAT-CAGGCT-CCT-GNAG
AA434486         35 TA-GTG-CCTT-CA--GGGTCC-GA-GG-AGCAT-CAGGCT-CCT-GCAG
N94336.RC        16 TAAGTG-CCTT-CAAGGGGTCC-GA-GGGAGCAT-CAAGNTTCCTTGCAG
2758102          17 TA-GTG-CCTT-CA--GGGTCC-GA-GG-AGCAT-CAGGCT-CCT-GCAG
AA434390.RC       1 TA-GTG-CCTT-CA--GGGTCC-GA-GG-AGCAT-CAGGCT-CCT-GCAG
1284649           1      G-CCTT-CA--GGGTCC-GA-GG-AGCAT-CAGGCT-CCT-GCAG
1979276           1                    GTCC-GA-GG-AGCAT-CAGGCT-CCT-GCAG
2325950           1                      CC-GA-GG-AGCAT-CAGGCT-CCT-GCAG
1554785           1                                                GCAG
                    .. ... .... ..  .+++.. .. .. ..+.. ...... ... ....
<consen01>     1574 TA GTG CCTT CA  GGGTCC GA GG AGCAT CAGGCT CCT GCAG

R15713          299 AGCCC-AT-CCCCCC-G-CC-ACACCCACACGGTGG-AGCT-TGCCT
H00901          280 AGCCCCATTCCCCCC-G-CC-ACANCCACACGGTGGGAGNTTGCCTCTTT
R67665          278 AGCCCCAT-CCCCCC-G-CC-ACACCCACACGGTGG-AGCTTGCCTCTT-
R73342          260 AGCCCCAT-CCCCCC-G-CC-ACACCCACACGGTGG-AGCT-GCCTCTT-
R73357          259 AGCCCCAT-CCCCCC-G-CC-ACACCCACACGGTGG-AGCT-GCCTCTT-
AA211365.RC     212 AGCCCCAT-CCCCCC-G-CC-ACACCCACACGGTGG-AGCT-GCCTCTT-
1734407         212 AGCCCCAT-CCCCCC-G-CC-ACAC
2620380         212 ANCCCCAN-TNCCNC-G-AN-ACACCCACACGGTGG-AGCT-GNNTCAT-
2131632         207 ACGGTGAN-TNCTTT-C-TT-CCTCTCCTGTNCAAT-ANAT-T
AA130587.RC     189 AGCCCCAT-CCCCCC-G-CC-ACACCCACACGGTGG-AGCT-GCCTCTT-
R72869.RC       173 AGCCCCAT-CCCCCC-G-CC-ANACCCANACGGTGG-AGCT-GCCTCTT-
237971          168 AGCCC-AT-CCCCCC-G-CC-ACACCCACACGGTGG-AGCT-GCCTCTT-
238502          162 AGCCC-AT-CCCCCC-G-NC-ACACCCACACGGTGG-AGCT-GCCTCTT-
R72008.RC       139 AGCCCCAT-CCCCCC-G-CCCANACCCACACGGTGG-AGCT-GCCTCTT-
```

FIGURE 4AA

```
AA480065.RC     124 AGCCCCAT-CCCCCC-G-CC-ACACCCACACGGTGG-AGCT-GCCTCTT-
1818182         114 AGCCCCAT-NCCCCC-G-CC-ANACCCACACGGTGG-AGCT-GCCTCTT-
2194762         114 AGCCCCAT-CCCCCC-G-CN-ACACCCACACGGTGG-AGCT-GCCTCTT-
1818702         114 AGCCCCAT-CNCCCC-G-CC-ACNCCCACACGGTGG-AGCT-GCCTCTT-
AA010655        114 AGCCC-AT-CCCCCC-G-CC-ACACCCACACGGTGG-AGCT-GCCTCTT-
AA010656.RC      92 AGCCCCAT-CCCCCC-G-CC-ACACCCACACGGTGG-AGCT-GCCTCTT-
R16077.RC        90 AGCCCCAT-CCCCCC-GNCC-AAACCCANANGGTGG-AGCT-GCCTCTT-
R93337.RC        89 AGCCCCAT-CCCCCC-G-CC-AAAACCANACGGTGG-AGCT-GCCTCTT-
2053751          83 AGCCCCAT-CCCCCC-G-CC-ACACCCACACGGTGG-AGCT-GCCTCTT-
H00902.RC        84 AGCCCCAT-NCCCCC-G-CC-AAACCAA-ANGGTGG-AGCT-GCCTCTT-
R72882.RC        80 AGCCCCAT-CCCCCC-G-CCCANACCCANACGGTGG-AGCT-GCCTCTT-
AA434486         74 AGCCCCAT-CCCCCC-G-CC-ACACCCACACGGTGG-AGCT-GCCTCTT-
N94336.RC        61 AGCCCCAT-CCCCCCCG-CC-ACACCCACACGGTGG-AGCT-GCCTCTT-
2758102          56 ANCCCCAT-CCCCCC-G-CC-ACACCCACACGGTGG-AGCT-GCCTCTT-
AA434390.RC      40 AGCCCCAT-CCCCCC-G-CC-ACACCCACACGGTGG-AGCT-GCCTCTT-
1284649          36 AGCCCCAT-CCCCCC-G-CC-ACACCCACACGGTGG-AGCT-GCCTCTT-
1979276          27 ANCCCCAT-CCCCCC-G-CC-ACACCCACACGGTGG-AGCT-GCCTCTT-
2325950          25 ANCCCCAT-CCCCCC-G-CC-ACACCCACACGGTGG-AGCT-GCCTCTT-
1554785           5 AGCCCCAT-CCCCCC-G-CC-ACACCCACACGGTGG-AGCT-GCCTCTT-
R66066.RC         1   CCCCAT-TCCCCC-GNCC-AAACCCAAANGGTGG-AGCT-GCCTCTT-
807796            1        AT-NCCNNC-G-GC-ANANCCACACGGTGG-AGCT-GNCTNTT-
630305            1               CC-G-CC-ACACCCACACGGTGG-AGCT-GCCTCTT-
3277068           1                                    GG-AGCT-GCCTCTT-
T40555.RC         1                                         CT-GCCTCTT-
                    +.....+. ..+... . .. ............... +..+ ......+
<consen01>     1613 AGCCCCAT CCCCCC G CC ACACCCACACGGTGG AGCT GCCTCTT
```

FIGURE 4AB

H00901	327	CCTTTCCCCT-TCTN-CCTGTTTGCCCATTAT-TTCAGGATTTTTGGGAT
R67665	322	CCTT-CCNTT-CCTC-CCTGTT-GCCCATACT-CAGCATCTTGGATGAAA
R73342	303	CCTT-CCCCT-CCTTGCCTGTT-GCCCATANT-CAGCATTTTGGATGAAA
R73357	302	CCTT-CCCCTTCCTC-CCTGTT-GCCCATAATTCAGCATCTTGGATGAAA
AA211365.RC	255	CCTT-CCCCT-CCTC-CCTGTT-GCCCATACT-CAGCATCTCGGATGAAA
2620380	255	CCTT-NCCCT-CCTC-CCTGTT-GCCCATA
AA130587.RC	232	CCTT-CCCCT-CNTN-CNTGTT-GCCCATACT-CAGCATCTCGGATGAAA
R72869.RC	216	CCTT-CCCCT-CCTC-CCTGTT-GCCCATACT-CAGCATCTCGGATGAAA
237971	210	CCT
238502	204	CCTT-CCCCT-CCTC-CCTGTT-GC
R72008.RC	183	CCTT-CCCCT-CCTC-CCTGTT-GCCCATACT-CAGCATCTCGGATGAAA
AA480065.RC	167	CCTT-CCCCT-CCTC-CCTGTT-GCCCATACT-CAGCATCTCGGATGAAA
1818182	157	CCTT-CCCCT-CCTC-CCTGTT-GCCCATACT-CAGCATCTCGGATGAAA
2194762	157	CCTT-CCCCT-CCTC-CCTGTT-GCCCATACT-CAGCATCTCGGATGAAA
1818702	157	CCTT-CCCCT-CCTC-CCTGTT-GCCCATA-T-CAGCATCTCGGATGAAA
AA010655	156	CCTT-CCCCT-CNTC-CCTGTT-GCCCATACT-CAGCATCTCGGATGAAA
AA010656.RC	135	CCTT-CCCCT-CCTN-CNTGTT-GCCCATACT-CAGCATCTCGGATGAAA
R16077.RC	134	CCTT-CCCCT-CCTC-CCTGTT-GCCCATACT-CAGCATCTCGGATGAAA
R93337.RC	132	CCTT-CCCCT-CCTC-CCTGTT-GCCCATACT-CAGCATCTCGGATGAAA
2053751	126	CCTT-CCCCT-CCTC-CCTGTT-GCCCATACT-CAGCATCTCGGATGAAA
H00902.RC	126	CCTT-CCCCT-CCTC-CCTGTT-GCCCATACT-CAGCATCTCGGATGAAA
R72882.RC	124	CCTT-CCCCT-CCTC-CCTGTT-GCCCATACT-CAGCATCTCGGATGAAA
AA434486	117	CCTT-CCCCT-CCTC-CCTGTT-GCCCATACT-CAGCATCTCGGATGAAA
N94336.RC	105	CCTT-CCCCT-CCTC-CCTGTT-GCCCATACT-CAGCATCTCGGATGAAA
2758102	99	CCTT-CCCCT-CCTC-CCTGTT-GCCCATACT-CAGCATCTCGGATGAAA
AA434390.RC	83	CCTT-CCCCT-CCTC-CCTGTT-GCCCATACT-CAGCATCTCGGATGAAA
1284649	79	CCTT-CCCCT-CCTC-CCTGTT-GCCCATACT-CAGCATCTCGGATGAAA
1979276	70	CCTT-CCCCT-CCTC-CCTGTT-GCCCATACT-CAGCATCTCGGATGAAA
2325950	68	CCTT-CCCCT-CCTC-CCTGTT-GCCCATACT-CAGCATCTCGGATGAAA
1554785	48	CCTT-CCCCT-CCTC-CCTGTT-GCCCATACT-CAGCATCTCGGATGAAA
R66066.RC	43	CCTT-CCCCT-CCTC-CCTGTT-GCCCATACT-CAGCATCTCGGATGAAA
807796	38	CCTT-CCCNT-NCTG-CCNGTT-GCCCATACT-CAGCATNTNGGATGAAA
630305	32	CCTT-CCCCT-CCTC-CCTGTT-GCCCATACT-CAGCATCTCGGATGAAA
3277068	14	CCTT-CCCCT-CCTC-CCTGTT-GCCCATACT-CAGCATCTCGGATGAAA
T40555.RC	10	CCTT-CCCCT-CCTC-CCTGTT-GCCCATACT-CAGCATCTCGGATGAAA

FIGURE 4AC

```
756679            1                          CCTGTT-GCCCATACT-NAGCATCTCGGATGAAA
AA047596.RC       1                              TT-GCCCATACT-CAGCATCTCGGATGAAA
3142901           1                                     ATACT-CAGCATCTCGGATGAAA
762960            1                                                CTCGGATGAAA
                    ++++ .+..+ ..+. +..+++ ++++++..+ .......+.....+.+.
<consen01>     1656 CCTT CCCCT CCTC CCTGTT GCCCATACT CAGCATCTCGGATGAAA

H00901          374 TGAAA-GGGGTTC-CTT-TGTTTC-TTNAGGG-TTTTCAA-G--G-GGA-
R67665          367 GGGGT-CCTTTGT-TCTTNAGGTT-TCAAGGG-AGGGGGG-T--TTGTT-
R73342          349 GGGGTTCCCTTGTTCCTTCAGGGTTNCAAGGGGAACGGGG-GGTT-GNTT
R73357          349 GGGNTTCCCTTGTTCCT-TAGGGTTCCAAGGGGAGGGGGG-G--TTGNT-
AA211365.RC     300 GGGCT-CCCTTGT-CCT-CANGNT-CCACGGG-AGCGGGG-C--T-GCT-
AA130587.RC     277 GGGCT-CCCTTGT-CCT-CANGNT-CCACGGG-AGCGGGG-C--T-GCT-
R72869.RC       261 GGGCT-CCCTTGT-CCT-CAGG-T-CCACGGG-AGCGGGG-C--T-GCT-
R72008.RC       228 GGGCT-CCCTTGT-CCT-CAGG-T-CCACGGG-AGCGGGG-C--T-GCT-
AA480065.RC     212 GGGCT-CCCTTGT-CCT-CAGGCT-CCACGGG-AGCGGGG-C--T-GCT-
1818182         202 GGGCT-CCCTTGT-CCT-CAGGCT-CCACGGG-AGCGGGG-C--T-GCT-
2194762         202 GGGCT-CCCTTGT-CCT-CAGGCT-CCACGGG-AGCGGGG-C--T-GCT-
1818702         201 GGGNT-CCCTTGT-C-T-NAGGNT-CCACGGG-A-CGGGG-T--T-GTT-
AA010655        201 GGGCT-CCCTTGT-CCT-CAGGCT-CCACGGG-AN-GGGG-C--T-GCT-
AA010656.RC     180 GGGCT-CCCTTGT-CCT-CNAGNT-CCACGGG-AGCGGGG-C--T-GCT-
R16077.RC       179 GGGCT-CCCTTGT-CCT-CANGNT-CCACGGG-ANCGGGG-C--T-GCT-
R93337.RC       177 GGGCT-CCCTTGT-CCT-CAGG-T-CCACGGG-ANCGGGG-C--T-GCT-
2053751         171 GGGCT-CCCTTGT-CCT-CAGGCT-CCACGGG-AGCGGGG-C--T-GCT-
H00902.RC       171 GGGCT-CCCTTGT-CCT-CAGG-T-CCANGGG-ANCGGGG-C--T-GCT-
R72882.RC       169 GGGCT-CCCTTGT-CCT-CAGG-T-CCACGGG-AGCGGGG-C--T-GCT-
AA434486        162 GGGCT-CCCTTGT-CCT-CAGGCT-CCACGGG-A-CGGGG-C--T-GCT-
N94336.RC       150 GGGCT-CCCTTGT-CCT-CAGG-T-CCACGGG-AGCGGGG-C--T-GCT-
2758102         144 GGGCT-CCCTTGT-CCT-CAGGCT-CCACGGG-AGCGGGG-C--T-GCT-
AA434390.RC     128 GGGCT-CCCTTGT-CCT-CAGGCT-CCACGGG-AGCGGGG-C--T-GCT-
1284649         124 GGGCT-CCCTTGT-CCT-CAGGCT-CCACGGG-AGCGGGG-C--T-GCT-
1979276         115 GGGCT-CCCTTGT-CCT-CAGGCT-CCACGGG-AGCGGGG-C--T-GCT-
2325950         113 GGGCT-CCCTTGT-CCT-CAGGCT-CCACGGG-AGCGGGG-C--T-GCT-
1554785          93 GGGCT-CCCTTGT-CCT-CAGGCT-CCACGGG-AGCGGGG-C--T-GCT-
R66066.RC        88 GGGCT-CCCTTNT-CCT-CAGG-T-CCACGGG-ANCGGGG-C--T-GCT-
```

FIGURE 4AD

```
807796           83 GGGNT-NCCTTGT-CCT-GAGGCT-NCACGGG-AGCGGGGC--T-GCT-
630305           77 GGGCT-CCCTTGT-CCT-CAGGCT-CCANGGG-AGCGGGG-C--T-GCT-
3277068          59 GGNNT-CCCTTGT-CCT-CAGGCT-CCACGGG-AGCNGGG-C--T-GCT-
T40555.RC        55 GGGCT-CCCTTGT-CCT-CAGCGT-CCACGGG-AGCGGGG-C--T-GCT-
756679           33 GGGNT-NCCTTGT-CCT-CAGGNT-TCANGGG-AGCGGGG-C--T-NCT-
AA047596.RC      29 GGGCT-CCCTTGT-CNT-CNAGNT-CCACGGG-AGCGGGG-C--T-GCT-
3142901          23 GGGCT-CCCTTGT-CCT-CAGGCT-CCACGGG-AGCGGGG-C--T-GCT-
762960           12 GGGCT-CCCTTGT-CCT-CAGGCT-CCACGGG-AGCGGGG-C--T-GCT-
2729902           1       CCCTTGT-CCT-CAGGCT-CCACGGG-AGCNGGG-C--T-GCT-
1978442           1       CCCTTGT-CCT-CAGGCT-CCACGNN-AGCGGGG-C--T-GCT-
2306960           1                                       GC-A--A-CCG-
1926134           1                                               GCT-
                    .+...  ....+.. ..+ ...... ....+.. ....... .  . ...
<consen01>     1701 GGGCT CCCTTGT CCT CAGGCT CCACGGG AGCGGGG C  T GCT

H00901          414 NC-GGGGGGTTNTT--GG-G-AGAAGAGC-GG-GGGGAATTTTCCAACAA
R67665          409 GGGAGAGAGCNGGG--GA-A-TT
R73342          397 GG-AGAGAGCGGGGGAAA-T-TCCCACCAACA-TTGGGGGNATTCCGGNA
R73357          394 GGGAGAGAGCGGGGGGAA-T-TNCCACCAACA-TTGGGGGCATT
AA211365.RC     340 GG-AGAGAGCGGGG--AA-C-TCCCACCA-CA-GTGGGGCATCCGGCACT
AA130587.RC     317 GG-AGAGAGCGGGG--AA-C-NTCCACCA-CA-GTGGGGCATCCGGCACT
R72869.RC       300 GG-AGAGAGCGGGG--AA-CCTCCCACCA-CA-GTGGGGCATCCGGCACT
R72008.RC       267 GG-AGAGAGCGGGG--AA-N-CTCNACCA-NA-GTGGGGGATCCGGNACT
AA480065.RC     252 GG-AGAGAGCGGGG--AA-C-TCCCACCA-CA-GTGGGGCATCCGGCACT
1818182         242 GG-AGAGAGCGGGG--AA---TCCCACC
2194762         242 GG-AGAGAGCGGGG--AA-C-TCCCACCA-CA-GTGGGGCATCCG
1818702         239 GG-AAAAA-CGGGG--AA-N-TCCNCCAA-AG-GGGGNTCGGNATTA
AA010655        240 GG-AGAGAGCGGGG--AA-C-TCCCACCA-CANGTGGGGCATCCGG-ACT
AA010656.RC     220 GG-AGAGAGCGGGG--AA-C-TCC-ACCA-CA-GTGGGGCATCCGGCACT
R16077.RC       219 GG-AGAGAGCGGGG--AAAC-TNC-ACCA-CA-GTGGGGCATCCGGCACT
R93337.RC       216 GG-AGAGAGCGGGG--AA-C-TTCCACCA-CA-GTGGGGCATCCGGCACT
2053751         211 GG-AGAGAGCGGGG--AA-C-TCCCACCA-CA-GTGGGGCATCCGGCACT
H00902.RC       210 GG-AGAGANCGGGG--NA-C-TCCCACCA-NA-GTGGGGCATCCGGNACT
R72882.RC       208 GG-AGAGAGCGGGG--AA-C-TCCCACCA-CA-GTGGGGCATCCGGCACT
AA434486        201 GG-AGAGAGCGGGG--AA-C-TCCCACCA-CA-GTGGGGCATCCGGCACT
```

FIGURE 4AE

```
N94336.RC          189 GG-AGAGAGCGGGG--AA-C-TTCCACCA-CA-GTGGGGCATCCGGCACT
2758102            184 GG-AGAGAGCGGGG--AA-C-TCCCACCA-CA-GTGGGGCATCCGGCACT
AA434390.RC        168 GG-AGAGAGCGGGG--AA-C-TCCCACCA-CA-GTGGGGCATCCGGCACT
1284649            164 GG-AGAGAGCGGGG--AA-C-TCCCACCA-CA-GTGGGGCATCCGGCACT
1979276            155 GG-AGAGAGCGGGG--AA-C-TCCCACCA-CA-GTGGGGCATCCGGCACT
2325950            153 GG-AGAGAGCGGGG--AA-C-TCCCACCA-CA-GTGGGGCATCCGGCACT
1554785            133 GG-AGAGAGCGGGG--AA-C-TCCCACCA-CA-GTGGGGCATCCGGCACT
R66066.RC          127 GG-AGAGANCGGGG--AA-C-TCCCACCA-NA-GTGGGGCATCCGGNACT
807796             124 GG-AGAGAGCGGGG--AA-C-TTCCACCA-CA-ATGGGGNATCCGGCANT
630305             117 GG-AGAGAGCGGGG--AA-C-TNCCANCA-CA-GTGGGGCATCCGGCACT
3277068             99 GG-AGAGAGCGGGG--AA-C-TCCCACCA-CA-GTGGGGCATCCGGCACT
T40555.RC           95 GG-AGAGAGCGGGG--AA-C-TCCCACCA-CA-GTGGGGCATCNGGCACT
756679              73 NG-AGAGAGCGGGG--AA-C-TTCCACCA-NA-NTGGGGCAATCGGCAAT
AA047596.RC         69 GG-AGAGAGCGGGG--AA-C-TCCCACCA-CA-GTGGGGCATCCGGCACT
3142901             63 GG-AGAGAGCGGGG--AA-C-TCCCACCA-CA-GTGGGGCATCCGGCACT
762960              52 GG-AGAGAGCGGGG--AA-C-TCCCACCA-CA-GTGGGGCATCCGGCACT
2729902             36 GG-AGAGAGCGGGG--AA-C-TCCCACCA-CA-GTGGGGCATCCGGCACT
1978442             36 GG-AGANAGCGGGG--AA-C-TCCCACCA-CA-GTGGGGCATCCGGCACT
2306960              8 CG-ACTTCAGCTGA--NG-C-AGCCTCAC-AG-CCTGCCATCTGGTGCCT
1926134              4 GG-AGANAGCGGGG--AA-C-TCCCACCA-CA-GTGGGGCATCCNGCACT
1737558              1           GGGG--AA-C-TCCCACCA-CA-GTGGGGCATCCGGCACT
1735126              1           GGGG--AA-C-TCCCACCA-CA-GTGTGTCATCCGGCACT
1929940              1                 C-TCCCACCA-CA-GTGGGGCATCCNGCACT
AA283209.RC          1                    CCCACCA-CA-GTGGGGCATCCGGCACT
                                  .. ..... .. ..... .. ................
<consen01>        1741 GG AGAGAGCGGGG  AA C TCCCACCA CA GTGGGGCATCCGGCACT

H00901             457 ATTGGGGGCATTTCGG
R73342             443 ATTGAAGGCTTGGTTTTTCT
AA211365.RC        383 GAAGCCCTGGTGTTCCTGGT-C-A-CGT-CNCCCAGGGGACCCTGCCCGC
AA130587.RC        360 GAAGCCCTGGTGTTCCTGGT-C-A-NTT-CCCCCAGGG-ACCCTGCCCCC
R72869.RC          344 GAAGCCCTGGTGTTCCTGGT-C-ANCGT-CCCC-AGGGGACCCTGCCCGC
R72008.RC          310 GAAGCCCTGNTGGNCCTGGTNC-A-CCG-CGCCCAGGG-ACCCTGCCCCC
AA480065.RC        295 GAAGCCCTGGTGTTCCTGGT-C-A-CGT-CCCCCAGGGGACCCTGCCCGC
AA010655           283 GAAGCCCTGGTGTTCCTGGT-C-A-CGT-CCCC-ANGGGACC
```

FIGURE 4AF

```
AA010656.RC    262 GAAGCCCTGGTGTTCCTGGT-CNA-CNT-CCCCCAGGG-ACCCTGCCCCC
R16077.RC      262 GAAGCCCTGGTGTTCCTGGT-C-N-AAC-NCCCCAGGGGACCCTGCCCAC
R93337.RC      259 GAAGCCCTGGTGTTCCTGGT-C-AACGT-CCCC-AGGGGACCCTGCCCAC
2053751        254 GAAGCCCTGGTGTTCCTGGT-C
H00902.RC      253 GAANCCCTGGTGTNCCTGGT-C-AACGT-CCCCCAGGGN-CCCTGCCCCC
R72882.RC      251 GAAGCCCTGGTGTTCCTGGT-C-ANCGT-CCCC-AGGGGACCCTGCCCGC
AA434486       244 GAAGCCCTGGTGTTCCTGGT-C-A-CGT-CCCCCAGGGGACCCTGCCCGC
N94336.RC      232 GAAGCCCTGGTGTTCCTGGT-CGA-CGT-CCCCCAGGG-ACCCTGCCCCC
2758102        227 GAAGCCCTGGTGTTCCTGGT-C-A-CGT-CCCCCAGGGGACCCT
AA434390.RC    211 GAAGCCCTGGTGTTCCTGGT-C-AACGT-CCCCCAGGGGACCCTGCCCGC
1284649        207 GAAGCCCTGGTGTTCCTGGT-C-A-CGT-CCCCCAGGGGACGC
1979276        198 GAAGCCCTGGTGTTCCTGGT-C-A-CG
2325950        196 GAAGCCCTGGTGTTCCTGGT-C-A-CGT-CCCCCAGGGGACCCTGCCCCT
1554785        176 GAA
R66066.RC      170 GAANCCCTGGTGTNCCTGGT-C-AACGT-CCCCCAGGGN-CCCTGCCCCC
807796         167 GAAGCCCTGGTGTTCCTGGT-N-A-AGT-CCCCNAGGGGACCNTGNCCGG
630305         160 GAAGCCCTGGTGTTC-TGGT-C-A-NGT-NCCCCANGGGACCCTGCCCCC
3277068        142 GAAGCCCTGGTGTTCCTGGT-C-A-CGT-CCCCCAGGGGACCCTGCCCGC
T40555.RC      138 GAAGCCCTGGTGTTCCTGGT-C-A-CGT-CCCCCAGGG-ACCCTGCCCCC
756679         116 GAAGGCCTGGTGTTTCTGGT-C-A-AGTTCCCCCAGGGGACCCTGCCCCC
AA047596.RC    112 GAAGCCCTGGTGTTCCTGGT-CTA-CTT-CCCCCNAGGGACCCTGCCCCC
3142901        106 GAAGCCCTGGTGTTCCTGGT-C-A-CGT-CCCCCAGGGGACCCTGCCCCC
762960          95 GAAGCCCTGGTGTTNCTGGT-C-A-AGT-CCCCCAGGGGACCCTGCCCCC
2729902         79 GAAGCCCTGGTGTTCCTGGT-C-A-CGT-CCCCCAGGGGACCCTGCCCCC
1978442         79 GAAGCCCTGGTGTNCCTGGN-C-A-CGT-CCCCCAGGGGACCCTGCCCCC
2306960         51 CCTGCCACCTGGTGCCTCTC-G-G-CTC-GGTGACAGCCAACCTGCCCCC
1926134         47 GAAGCCCTGGTGTTCCTGGT-C-A-CGT-CCCCCAGGGGACCCTGCCCCC
1737558         35 GAAGCCCTGGTGTTCCTGGT-C-A-CGT-CCCCCAGGGGACCCTGCCCCC
1735126         35 GAAGCCCTGGTGTTCCTGGT-C-A-CGT-CCCCCAGGGGACCCTGCCCCC
1929940         29 GAAGCCCTGGTGTTCCTGGT-C-A-CGT-CCCCCAGGGGACCCTGCCCCC
AA283209.RC     27 GAAGCCCTGGTGTTCCTGGT-C-A-CGT-CCCCCAGGGGACCCTGCCCGC
2126005          1                                GGACCCTGCCCCC
                   ...............+... . . ... .......+..+...++.++ .
<consen01>    1784 GAAGCCCTGGTGTTCCTGGT C A CGT CCCCCAGGGGACCCTGCCCGC
```

FIGURE 4AG

```
AA211365.RC    429 TTCCTGGACTTCGTGCCTTACTGAGTCTCTAAGACTTTTTCTAATAAACA
AA130587.RC    405 TTCCTGGACTTCGTGCCTTACTGAGTCTCTAAGACTTTTTCTAATAAACA
R72869.RC      390 TTCCTGGACTTCGTGCCTTACTGAGTCTCTAAGACTTTTTCTAATAAACA
R72008.RC      356 NNCCTGGACTTCGTGCCTTACTGAGTCTCTAAGACTTTTTCNAATAAAAA
AA480065.RC    341 TTCCTGGACTTCGTGCCTTACTGAGTCTCTAAGACTTTTTCTAATAAACA
AA010656.RC    308 TTCCTGGACTTCGTGCCTTACTGAGTCTCTAAGACTTTTTCTAATAAACA
R16077.RC      308 TTCCTGGACTTCGTGCCTTACTGAGTCTCTAAGACTTTTTCTAATAAACA
R93337.RC      305 TTCCTGGACTTCGTGCCTTACTGAGTCTCTAAGACTTTTTCTAATAAACA
H00902.RC      299 TNCCTGGNCTTCGTGCCTTACTGAGTCTCTAAGNCTTTTTCTAATAAACA
R72882.RC      297 TTCCTGGACTTCGTGCCTTACTGAGTCTCTAAGACTTTTTCTAATAAACA
AA434486       290 TTCCTGGACTTCGTGCCTTACTGAGTCTCTAAGACTTTTTCTAATAAACA
N94336.RC      278 TTCCTGGA-TTCGTGCCTTACTGAGTCTCTAAGACTTTTTCTAATAAACA
AA434390.RC    258 TTCCTGGACTTCGTGCCTTACTGAGTCTCTAAGACTTTTTCTAATAAACA
2325950        242 TTCCTGGACTTCGTGCCTTATNGAGTCTCNA
R66066.RC      216 TNCCTGGNCTTCGTGCCTTACTGAGTCTCTNAGACTTTTTCTAATAAACA
807796         213 TTGNTGGAT
630305         205 TTCCTGGACTTNGTGCCTTANTGAGTCT-TAAGACTTTTTCTAATAAACA
3277068        188 TTCCTGGACTTCGTGCCTTACTGAGTCTCTAAGACTTTTTCTAATAAACA
T40555.RC      183 TTCCTGGACTTCGTGCCTTACTGAGTCTCTAAGACTTTTTCTAATAAACA
756679         163 CTTTCCTGGGACTTTGGTGGNCCTTAATTGAGGTCTTTTAAGGACTTTTT
AA047596.RC    159 TTNCTGGACTTCGTGCCTTACTGAGTCTCTAAGACTTTTTCTAATAAACA
3142901        152 TTCCTGGACTTCGTGCCTTACTGAGTCTCTAAGACTTTTTCTAATAAACA
762960         141 TTCCTGGACTNCGTGCCTTACTGAGTCTCTAAGACTTTTTCTAATAAACA
2729902        125 TTCCTGGACTTCGTGCCTTACTGAGTCTCTAAGACTTTTTCTAATAAACA
1978442        125 TTCCNGGACTNCGTGCCTTACTGAGTCTCTAAGACTTTTNCTAATAAACA
2306960         97 TTCCTGGACTTCGTGCCTTACTGAGTCTCTAAGACTTTTTCTAATAAACA
1926134         93 TTCCTGGACTTCGTGCCTTACTGAGTCTCTAAGACTTTTTCTAATAAACA
1737558         81 TTCCTGGACTTCGTGCCTTACTGAGTCTCTAAGACTTTTTCTAATAAACA
1735126         81 TTCCTGGACTTCGTGCCTTACTGAGTCTCTAAGACTT
1929940         75 TTCCTGGACTTCGTGCCTTACTGAGTCTCTAAGACTTTTTCTAATAAACA
AA283209.RC     73 TTCCTGGACTTCGTGCCTTACTGAGTCTCTAAGACTTTTTCTAATAAACA
2126005         14 TTCCTGGACTTCGTGCCTTACTGAGTCTCTAAGACTTTTTCTAATAAACA
                   .............+...+.........+....+.+++++....+......
<consen01>    1830 TTCCTGGACTTCGTGCCTTACTGAGTCTCTAAGACTTTTTCTAATAAACA
```

FIGURE 4AH

```
AA211365.RC      479 AGCCAGTGCGTGTAA
AA130587.RC      455 AGCCAGTGCGTGTACCATGAAAA
R72869.RC        440 AGNNAGTNCNTGTAAANAAAN
R72008.RC        406 NNTTTNAGTGCGTGTNAAATT
AA480065.RC      391 AGCCAGTGCGTGTACCATGAAAA
AA010656.RC      358 AGCCAGTGCGTGTACCN
R16077.RC        358 AGCCAGTGCGTGTACCAAAAAAA
R93337.RC        355 AGCCAGTGCGTGTTCANATNA
H00902.RC        349 AGCCAGT
R72882.RC        347 NG
AA434486         340 AGCCAGTGCGTGTACCA
N94336.RC        327 AGCCAGTGCGTGTAAAAAAAAA
AA434390.RC      308 AGCCAGTGCGTGTACCAAAAAAAAA
R66066.RC        266 AGCCAGTNCGTGTNCCAAAAANT
630305           254 AGNCANTGCGT
3277068          238 AGCCAGTGCGTGT
T40555.RC        233 AGCCAGTGCGTGT
756679           213 TTGTTAATTAAAACCAAAGCCCAAGTGGCGGTTGTAAAAAAAAANNAANN
AA047596.RC      209 AGCCAGTGCNTGTACC
3142901          202 AGCCAGTGCGTGTAAAA
762960           191 AGCCAGTGCGTGT
2729902          175 AGCCAGTGCGTGT
1978442          175 AGCCAGTGCGTGTAA
2306960          147 AGCCAGTGCGTGTACCATG
1926134          143 AGCCAGTGCGTGTAC
1737558          131 AGCCAGTGCGTGTAC
1929940          125 AGCCAGTGCGTGTAAAA
AA283209.RC      123 AGCCAGTGCGTGTACCATGAAAA
2126005           64 AGCCAGTGCGANTACCA
                     .................  ....eeeeeeeeeeeeeeeeeeeeeeeeeee
<consen01>      1880 AGCCAGTGCGTGTACCATGAAAAGTGGCGGTTGTAAAAAAAAANNAANN

756679           263 ANA
                     eee
<consen01>      1930 ANA
```

PROVISIONAL APPLICATION COVER SHEET

A/PROV

This is a request for filing a PROVISIONAL APPLICATION under 37 CFR 1.53(b)(2)

"EXPRESS MAIL" MAILING LABEL
Number: EM 370 628 341 US Date of Deposit April 29, 1998
I hereby certify that this paper or fee is being deposited with the United States Postal Service "Express Mail Post Office to Addressee" service under 37 C.F.R. § 10 on the date indicated above and is addressed to the Assistant Commissioner for Patents, Washington, D.C. 20231.
Typed or Printed Name: Kevin Crandell
Signed: K Green

Docket	P-66091/WHD/DAV	

INVENTOR(s)/APPLICANT(s)

LAST NAME	FIRST NAME	MIDDLE INITIAL	RESIDENCE (CITY AND EITHER STATE OR FOREIGN COUNTRY)
WOOD	WILLIAM	I.	San Mateo, California

TITLE OF THE INVENTION (280 characters max)

NOVEL MULTI-TRANSMEMBRANE POLYPEPTIDES HAVING SEQUENCE IDENTITY WITH MEMBRANE REGULATOR PROTEINS AND NUCLEIC ACIDS

CORRESPONDENCE ADDRESS

Walter H. Dreger
FLEHR HOHBACH TEST ALBRITTON & HERBERT LLP, 4 Embarcadero Center, Suite 3400, San Francisco

STATE	CA	ZIP CODE	94111	COUNTRY	US

ENCLOSED APPLICATION PARTS (check all that apply)

X	Specification	Number of Pages	41		Small Entity Statement
X	Drawings	Number of Sheets	6		Other (specify):

METHOD OF PAYMENT OF FILING FEES FOR THIS PROVISIONAL APPLICATION FOR PATENT (check one)

X	A check or money order is enclosed to cover the Provisional filing fees	PROVISIONAL FILING FEE AMOUNT ($)	$150
X	The Commissioner is hereby authorized to charge any underpayment or overpayment to Deposit Account Number: 06-1300 (Order No. P-66091/WHD/DAV)		

The invention was made by an agency of the United States Government or under a contract with an agency of the United States Government.

X No.

☐ Yes, the name of the U.S. Government Agency and the Government contract number are: _____

Respectfully submitted, Reg. # 39054

SIGNATURE: Walter H. Dreger Date: April 29, 1998

TYPED or PRINTED NAME: Walter H. Dreger REGISTRATION NO. (if appropriate): 24,190

☐ Additional inventors are being named on separately numbered sheets attached hereto.

USE ONLY FOR FILING A PROVISIONAL APPLICATION FOR PATENT

Burden Hour Statement: This form is estimated to take .2 hours to complete. Times will vary depending upon the needs of the individual case. Any comments on the amount of time you are required to complete this form should be sent to the Chief Information Officer, Assistant Commissioner for Patents, Washington, DC 20231. DO NOT SEND FEES OR COMPLETED FORMS TO THIS ADDRESS. SEND TO: Box Provisional Application, Assistant Commissioner for Patents, Washington, D.C. 20231.

583498

PATENT
Attorney Docket No.: P-66091/WHD/DAV
(Genentech Docket No.:PR1335)

NOVEL MULTI-TRANSMEMBRANE POLYPEPTIDES HAVING SEQUENCE IDENTITY WITH MEMBRANE REGULATOR PROTEINS AND NUCLEIC ACIDS ENCODING THE SAME

FIELD OF THE INVENTION

The present invention relates generally to the identification and isolation of novel DNA and to the recombinant production of novel multi-transmembrane polypeptides having sequence identity with membrane regulator proteins, designated herein as "PRO218" polypeptides.

BACKGROUND OF THE INVENTION

Membrane-bound proteins and receptors can play an important role in the formation, differentiation and maintenance of multicellular organisms. The fate of many individual cells, e.g., proliferation, migration, differentiation, or interaction with other cells, is typically governed by information received from other cells and/or the immediate environment. This information is often transmitted by secreted polypeptides (for instance, mitogenic factors, survival factors, cytotoxic factors, differentiation factors, neuropeptides, and hormones) which are, in turn, received and interpreted by diverse cell receptors or membrane-bound proteins. Such membrane-bound proteins and cell receptors include, but are not limited to, cytokine receptors, receptor kinases, receptor phosphatases, receptors involved in cell-cell interactions, and cellular adhesin molecules like selectins and integrins. For instance, transduction of signals that regulate cell growth and differentiation is regulated in part by phosphorylation of various cellular proteins. Protein tyrosine kinases, enzymes that catalyze that process, can also act as growth factor receptors. Examples include fibroblast growth factor receptor and nerve growth factor receptor.

Membrane-bound proteins and receptor molecules have various industrial applications, including as pharmaceutical and diagnostic agents. Receptor immunoadhesins, for instance, can be employed as therapeutic agents to block receptor-ligand interaction. The membrane-bound proteins can also be employed for screening of potential peptide or small molecule inhibitors of the relevant receptor/ligand interaction.

Efforts are being undertaken by both industry and academia to identify new, native

membrane bound proteins, particularly those having sequence identity with membrane regulator proteins. Many efforts are focused on the screening of mammalian recombinant DNA libraries to identify the coding sequences for novel receptor proteins.

SUMMARY OF THE INVENTION

Applicants have identified a cDNA clone that encodes a novel multi-transmembrane protein having sequence identity with membrane regulator proteins, wherein the polypeptide is designated in the present application as "PRO218".

In one embodiment, the invention provides an isolated nucleic acid molecule comprising DNA encoding a PRO218 polypeptide. In one aspect, the isolated nucleic acid comprises DNA encoding the PRO218 polypeptide having amino acid residues 1 through 455 of Figure 2 (SEQ ID NO:2), or is complementary to such encoding nucleic acid sequence, and remains stably bound to it under at least moderate, and optionally, under high stringency conditions. The isolated nucleic acid sequence may comprise the cDNA insert of the vector deposited on 28 April 1998 with the ATCC as DNA30867-1335 which includes the nucleotide sequence encoding PRO218.

In another embodiment, the invention provides a vector comprising DNA encoding a PRO218 polypeptide. A host cell comprising such a vector is also provided. By way of example, the host cells may be CHO cells, *E. coli*, or yeast. A process for producing PRO218 polypeptides is further provided and comprises culturing host cells under conditions suitable for expression of PRO218 and recovering PRO218 from the cell culture.

In another embodiment, the invention provides isolated PRO218 polypeptide. In particular, the invention provides isolated native sequence PRO218 polypeptide, which in one embodiment, includes an amino acid sequence comprising residues 1 through 455 of Figure 2 (SEQ ID NO:2). Optionally, the PRO218 polypeptide is obtained or is obtainable by expressing the polypeptide encoded by the cDNA insert of the vector deposited on 28 April 1998 with the ATCC as DNA30867-1335.

In another embodiment, the invention provides chimeric molecules comprising a PRO218 polypeptide fused to a heterologous polypeptide or amino acid sequence. An example of such a chimeric molecule comprises a PRO218 polypeptide fused to an epitope tag sequence or a Fc region of an immunoglobulin.

In another embodiment, the invention provides an antibody which specifically binds to a PRO218 polypeptide. Optionally, the antibody is a monoclonal antibody.

In another embodiment, the invention provides expressed sequence tags (ESTs)

comprising the nucleotide sequence of SEQ ID NO:3 or 4.

BRIEF DESCRIPTION OF THE DRAWINGS

Figures 1A and 1B show a nucleotide sequence (SEQ ID NO:1) containing the nucleotide sequence of a native sequence PRO218 cDNA (nucleotides 150-1514), wherein the nucleotide sequence (SEQ ID NO:1) is a clone designated herein as "UNQ192" and/or "DNA30867-1335". Also presented is the position of the initiator methionine residue, circled, and the stop codon, circled. The putative transmembrane domains of the protein are encoded by nucleotides beginning at nucleotides 258, 390, 597, 1011, 1161, 1272, 1422 in the figure. The complementary sequence and deduced amino acid sequence are also shown.

Figures 2A and 2B show amino acid sequences. Figure 2A shows the amino acid sequence (SEQ ID NO:2) derived from nucleotides 150-1514 of the nucleotide sequence shown in Figures 1A and 1B. Also shown in Figure 2B are a signal peptide, potential transmembrane domains, N-glycosylation sites, and a eukaryotic cobalamin-binding protein.

Figure 3 shows two novel ESTs, SEQ ID NOS:3 and 4.

Figure 4 shows from DNA or consensus DNA formed on the basis of SEQ ID NOS:3 and 4.

DETAILED DESCRIPTION OF THE PREFERRED EMBODIMENTS

I. Definitions

The terms "PRO218 polypeptide" and "PRO218" when used herein encompass native sequence PRO218 and PRO218 polypeptide variants (which are further defined herein). The PRO218 polypeptides may be isolated from a variety of sources, such as from human tissue types or from another source, or prepared by recombinant or synthetic methods.

A "native sequence PRO218 polypeptide" comprises a polypeptide having the same amino acid sequence as a PRO218 polypeptide derived from nature. Such native sequence PRO218 polypeptide can be isolated from nature or can be produced by recombinant or synthetic means. The term "native sequence PRO218 polypeptide" specifically encompasses naturally-occurring truncated or secreted forms of a PRO218 polypeptide (*e.g.*, soluble forms containing for instance, an extracellular domain sequence), naturally-occurring variant forms (*e.g.*, alternatively spliced forms) and naturally-occurring allelic variants of a PRO218 polypeptide. In one embodiment of the invention, the native sequence PRO218 polypeptide is a full-length or mature native sequence PRO218 polypeptide comprising amino acids 1 or 24 through 455 of SEQ ID NO:2. In another embodiment of the invention, the native

sequence PRO218 polypeptide is an extracellular domain of the full-length PRO218 protein. Optionally, the PRO218 polypeptide is obtained or obtainable by expressing the polypeptide encoded by the cDNA insert of the vector DNA30867-1335 deposited on 28 April 1998 with the ATCC.

The "PRO218 extracellular domain" or "PRO218 ECD" refers to a form of the PRO218 polypeptide which is essentially free of the transmembrane and cytoplasmic domains of the PRO218 polypeptide. Ordinarily, PRO218 ECD will have less than 1% of such transmembrane and/or cytoplasmic domains and preferably, will have less than 0.5% of such domains. It will be understood that any transmembrane domain identified for the PRO218 polypeptide of the present invention is identified pursuant to criteria routinely employed in the art for identifying that type of hydrophobic domain. The exact boundaries of a transmembrane domain may vary but most likely by no more than about 5 amino acids at either end of the domain as initially identified. .

"PRO218 variant" means an active PRO218 polypeptide as defined below having at least about 80% amino acid sequence identity with the PRO218 polypeptide having the deduced amino acid sequence shown in Figure 2A (SEQ ID NO:2) for a full-length or mature native sequence PRO218 polypeptide. Such PRO218 polypeptide variants include, for instance, PRO218 polypeptides wherein one or more amino acid residues are added, substituted or deleted, at the N- or C-terminus or within the sequence of Figure 2 (SEQ ID NO:2). Ordinarily, a PRO218 polypeptide variant will have at least about 80% amino acid sequence identity, preferably at least about 85% amino acid sequence identity, more preferably at least about 90% amino acid sequence identity and even more preferably at least about 95% amino acid sequence identity with the amino acid sequence of Figure 2 (SEQ ID NO:2), with or without the signal peptide. The variants provided herein exclude the polypeptides and nucleic acids described herein which the novel polypeptides have identity with and which are already known in the art.

"Percent (%) amino acid sequence identity" with respect to the PRO218 amino acid sequences identified herein is defined as the percentage of amino acid residues in a candidate sequence that are identical with the amino acid residues in a PRO218 polypeptide sequence, after aligning the sequences and introducing gaps, if necessary, to achieve the maximum percent sequence identity, and not considering any conservative substitutions as part of the sequence identity. Alignment for purposes of determining percent amino acid sequence identity can be achieved in various ways that are within the skill in the art, for instance, using publicly available computer software such as ALIGN or Megalign (DNASTAR) software.

Those skilled in the art can determine appropriate parameters for measuring alignment, including any algorithms needed to achieve maximal alignment over the full length of the sequences being compared. Similarity can be determined using the same programs.

"Percent (%) nucleic acid sequence identity" with respect to the PRO218 sequence identified herein is defined as the percentage of nucleotides in a candidate sequence that are identical with the nucleotides in the PRO218 sequence, after aligning the sequences and introducing gaps, if necessary, to achieve the maximum percent sequence identity. Alignment for purposes of determining percent nucleic acid sequence identity can be achieved in various ways that are within the skill in the art, for instance, using publicly available computer software such as ALIGN or Megalign (DNASTAR) software. Those skilled in the art can determine appropriate parameters for measuring alignment, including any algorithms needed to achieve maximal alignment over the full length of the sequences being compared. Similarity can be determined using the same programs.

The term "epitope tagged" where used herein refers to a chimeric polypeptide comprising a PRO218 polypeptide, or domain sequence thereof, fused to a "tag polypeptide". The tag polypeptide has enough residues to provide an epitope against which an antibody may be made, or which can be identified by some other agent, yet is short enough such that it does not interfere with the activity of the PRO218 polypeptide. The tag polypeptide preferably is also fairly unique so that the antibody does not substantially cross-react with other epitopes. Suitable tag polypeptides generally have at least six amino acid residues and usually between about 8 to about 50 amino acid residues (preferably, between about 10 to about 20 residues).

"Isolated," when used to describe the various polypeptides disclosed herein, means polypeptide that has been identified and separated and/or recovered from a component of its natural environment. Contaminant components of its natural environment are materials that would typically interfere with diagnostic or therapeutic uses for the polypeptide, and may include enzymes, hormones, and other proteinaceous or non-proteinaceous solutes. In preferred embodiments, the polypeptide will be purified (1) to a degree sufficient to obtain at least 15 residues of N-terminal or internal amino acid sequence by use of a spinning cup sequenator, or (2) to homogeneity by SDS-PAGE under non-reducing or reducing conditions using Coomassie blue or, preferably, silver stain. Isolated polypeptide includes polypeptide *in situ* within recombinant cells, since at least one component of the PRO218 polypeptide natural environment will not be present. Ordinarily, however, isolated polypeptide will be prepared by at least one purification step.

An "isolated" PRO218 polypeptide-encoding nucleic acid molecule is a nucleic acid

molecule that is identified and separated from at least one contaminant nucleic acid molecule with which it is ordinarily associated in the natural source of the PRO218 polypeptide-encoding nucleic acid. An isolated PRO218 polypeptide-encoding nucleic acid molecule is other than in the form or setting in which it is found in nature. Isolated PRO218 polypeptide-encoding nucleic acid molecules therefore are distinguished from the PRO218 polypeptide-encoding nucleic acid molecule as it exists in natural cells. However, an isolated PRO218 polypeptide-encoding nucleic acid molecule includes PRO218 polypeptide-encoding nucleic acid molecules contained in cells that ordinarily express PRO218 polypeptide where, for example, the nucleic acid molecule is in a chromosomal location different from that of natural cells.

The term "control sequences" refers to DNA sequences necessary for the expression of an operably linked coding sequence in a particular host organism. The control sequences that are suitable for prokaryotes, for example, include a promoter, optionally an operator sequence, and a ribosome binding site. Eukaryotic cells are known to utilize promoters, polyadenylation signals, and enhancers.

Nucleic acid is "operably linked" when it is placed into a functional relationship with another nucleic acid sequence. For example, DNA for a presequence or secretory leader is operably linked to DNA for a polypeptide if it is expressed as a preprotein that participates in the secretion of the polypeptide; a promoter or enhancer is operably linked to a coding sequence if it affects the transcription of the sequence; or a ribosome binding site is operably linked to a coding sequence if it is positioned so as to facilitate translation. Generally, "operably linked" means that the DNA sequences being linked are contiguous, and, in the case of a secretory leader, contiguous and in reading phase. However, enhancers do not have to be contiguous. Linking is accomplished by ligation at convenient restriction sites. If such sites do not exist, the synthetic oligonucleotide adaptors or linkers are used in accordance with conventional practice.

The term "antibody" is used in the broadest sense and specifically covers single anti-PRO218 polypeptide monoclonal antibodies (including agonist, antagonist, and neutralizing antibodies) and anti-PRO218 antibody compositions with polyepitopic specificity. The term "monoclonal antibody" as used herein refers to an antibody obtained from a population of substantially homogeneous antibodies, *i.e.*, the individual antibodies comprising the population are identical except for possible naturally-occurring mutations that may be present in minor amounts.

"Active" or "activity" for the purposes herein refers to form(s) of PRO218 which retain

the biologic and/or immunologic activities of native or naturally-occurring PRO218 polypeptide.

The terms "treating", "treatment" and "therapy" as used herein refer to curative therapy, prophylactic therapy, and preventative therapy.

The term "mammal" as used herein refers to any mammal classified as a mammal, including humans, cows, horses, dogs and cats. In a preferred embodiment of the invention, the mammal is a human.

II. Compositions and Methods of the Invention

A. Full-length PRO218 Polypeptide

The present invention provides newly identified and isolated nucleotide sequences encoding polypeptides referred to in the present application as PRO218. In particular, Applicants have identified and isolated cDNA encoding a PRO218 polypeptide, as disclosed in further detail in the Examples below. The PRO218-encoding clone was isolated from a human fetal kidney library. To Applicants present knowledge, the UNQ192 (DNA30867-1335) nucleotide sequence encodes a novel factor; using BLAST and FastA sequence alignment computer programs, no significant sequence identities to any known proteins were revealed. Some sequence identity was found with membrane regulator proteins, indicating that PRO218 may function as a membrane regulator.

B. PRO218 Variants

In addition to the full-length native sequence PRO218 polypeptide described herein, it is contemplated that PRO218 variants can be prepared. PRO218 variants can be prepared by introducing appropriate nucleotide changes into the PRO218-encoding DNA, or by synthesis of the desired PRO218 polypeptide. Those skilled in the art will appreciate that amino acid changes may alter post-translational processes of the PRO218 polypeptide, such as changing the number or position of glycosylation sites or altering the membrane anchoring characteristics.

Variations in the native full-length sequence PRO218 or in various domains of the PRO218 polypeptide described herein, can be made, for example, using any of the techniques and guidelines for conservative and non-conservative mutations set forth, for instance, in U.S. Patent No. 5,364,934. Variations may be a substitution, deletion or insertion of one or more codons encoding the PRO218 polypeptide that results in a change in the amino acid sequence of the PRO218 polypeptide as compared with the native sequence PRO218. Optionally the

variation is by substitution of at least one amino acid with any other amino acid in one or more of the domains of the PRO218 polypeptide. Guidance in determining which amino acid residue may be inserted, substituted or deleted without adversely affecting the desired activity may be found by comparing the sequence of the PRO218 polypeptide with that of homologous known protein molecules and minimizing the number of amino acid sequence changes made in regions of high homology. Amino acid substitutions can be the result of replacing one amino acid with another amino acid having similar structural and/or chemical properties, such as the replacement of a leucine with a serine, i.e., conservative amino acid replacements. Insertions or deletions may optionally be in the range of 1 to 5 amino acids. The variation allowed may be determined by systematically making insertions, deletions or substitutions of amino acids in the sequence and testing the resulting variants for activity in any of the *in vitro* assays described in the Examples below.

The variations can be made using methods known in the art such as oligonucleotide-mediated (site-directed) mutagenesis, alanine scanning, and PCR mutagenesis. Site-directed mutagenesis [Carter et al., Nucl. Acids Res., 13:4331 (1986); Zoller et al., Nucl. Acids Res., 10:6487 (1987)], cassette mutagenesis [Wells et al., Gene, 34:315 (1985)], restriction selection mutagenesis [Wells et al., Philos. Trans. R. Soc. London SerA, 317:415 (1986)] or other known techniques can be performed on the cloned DNA to produce the PRO218-encoding variant DNA.

Scanning amino acid analysis can also be employed to identify one or more amino acids along a contiguous sequence. Among the preferred scanning amino acids are relatively small, neutral amino acids. Such amino acids include alanine, glycine, serine, and cysteine. Alanine is typically a preferred scanning amino acid among this group because it eliminates the side-chain beyond the beta-carbon and is less likely to alter the main-chain conformation of the variant. Alanine is also typically preferred because it is the most common amino acid. Further, it is frequently found in both buried and exposed positions [Creighton, The Proteins, (W.H. Freeman & Co., N.Y.); Chothia, J. Mol. Biol., 150:1 (1976)]. If alanine substitution does not yield adequate amounts of variant, an isoteric amino acid can be used.

C. Modifications of PRO218

Covalent modifications of PRO218 polypeptides are included within the scope of this invention. One type of covalent modification includes reacting targeted amino acid residues of a PRO218 polypeptide with an organic derivatizing agent that is capable of reacting with selected side chains or the N- or C- terminal residues of a PRO218 polypeptide.

Derivatization with bifunctional agents is useful, for instance, for crosslinking PRO218 to a water-insoluble support matrix or surface for use in the method for purifying anti-PRO218 antibodies, and vice-versa. Commonly used crosslinking agents include, *e.g.*, 1,1-bis(diazo-acetyl)-2-phenylethane, glutaraldehyde, N-hydroxysuccinimide esters, for example, esters with 4-azidosalicylic acid, homobifunctional imidoesters, including disuccinimidyl esters such as 3,3'-dithiobis(succinimidylpropionate), bifunctional maleimides such as bis-N-maleimido-1,8-octane and agents such as methyl-3-[(p-azidophenyl)dithio]propioimidate.

Other modifications include deamidation of glutaminyl and asparaginyl residues to the corresponding glutamyl and aspartyl residues, respectively, hydroxylation of proline and lysine, phosphorylation of hydroxyl groups of seryl or threonyl residues, methylation of the α-amino groups of lysine, arginine, and histidine side chains [T.E. Creighton, Proteins: Structure and Molecular Properties, W.H. Freeman & Co., San Francisco, pp. 79-86 (1983)], acetylation of the N-terminal amine, and amidation of any C-terminal carboxyl group.

Another type of covalent modification of the PRO218 polypeptide included within the scope of this invention comprises altering the native glycosylation pattern of the polypeptide. "Altering the native glycosylation pattern" is intended for purposes herein to mean deleting one or more carbohydrate moieties found in native sequence PRO218 polypeptide, and/or adding one or more glycosylation sites that are not present in the native sequence PRO218 polypeptide.

Addition of glycosylation sites to PRO218 polypeptides may be accomplished by altering the amino acid sequence thereof. The alteration may be made, for example, by the addition of, or substitution by, one or more serine or threonine residues to the native sequence PRO218 polypeptide (for O-linked glycosylation sites). The PRO218 amino acid sequence may optionally be altered through changes at the DNA level, particularly by mutating the DNA encoding the PRO218 polypeptide at preselected bases such that codons are generated that will translate into the desired amino acids.

Another means of increasing the number of carbohydrate moieties on the PRO218 polypeptide is by chemical or enzymatic coupling of glycosides to the polypeptide. Such methods are described in the art, e.g., in WO 87/05330 published 11 September 1987, and in Aplin and Wriston, CRC Crit. Rev. Biochem., pp. 259-306 (1981).

Removal of carbohydrate moieties present on the PRO218 polypeptide may be accomplished chemically or enzymatically or by mutational substitution of codons encoding for amino acid residues that serve as targets for glycosylation. Chemical deglycosylation techniques are known in the art and described, for instance, by Hakimuddin, et al., Arch.

Biochem. Biophys., 259:52 (1987) and by Edge et al., Anal. Biochem., 118:131 (1981). Enzymatic cleavage of carbohydrate moieties on polypeptides can be achieved by the use of a variety of endo- and exo-glycosidases as described by Thotakura et al., Meth. Enzymol., 138:350 (1987).

Another type of covalent modification of PRO218 comprises linking the PRO218 polypeptide to one of a variety of nonproteinaceous polymers, *e.g.*, polyethylene glycol, polypropylene glycol, or polyoxyalkylenes, in the manner set forth in U.S. Patent Nos. 4,640,835; 4,496,689; 4,301,144; 4,670,417; 4,791,192 or 4,179,337.

PRO218 polypeptides of the present invention may also be modified in a way to form chimeric molecules comprising a PRO218 polypeptide fused to another, heterologous polypeptide or amino acid sequence. In one embodiment, such a chimeric molecule comprises a fusion of a PRO218 polypeptide with a tag polypeptide which provides an epitope to which an anti-tag antibody can selectively bind. The epitope tag is generally placed at the amino- or carboxyl- terminus of the PRO218 polypeptide. The presence of such epitope-tagged forms of a PRO218 polypeptide can be detected using an antibody against the tag polypeptide. Also, provision of the epitope tag enables the PRO218 polypeptide to be readily purified by affinity purification using an anti-tag antibody or another type of affinity matrix that binds to the epitope tag. In an alternative embodiment, the chimeric molecule may comprise a fusion of a PRO218 polypeptide with an immunoglobulin or a particular region of an immunoglobulin. For a bivalent form of the chimeric molecule, such a fusion could be to the Fc region of an IgG molecule.

Various tag polypeptides and their respective antibodies are well known in the art. Examples include poly-histidine (poly-his) or poly-histidine-glycine (poly-his-gly) tags; the flu HA tag polypeptide and its antibody 12CA5 [Field et al., Mol. Cell. Biol., 8:2159-2165 (1988)]; the c-myc tag and the 8F9, 3C7, 6E10, G4, B7 and 9E10 antibodies thereto [Evan et al., Molecular and Cellular Biology, 5:3610-3616 (1985)]; and the Herpes Simplex virus glycoprotein D (gD) tag and its antibody [Paborsky et al., Protein Engineering, 3(6):547-553 (1990)]. Other tag polypeptides include the Flag-peptide [Hopp et al., BioTechnology, 6:1204-1210 (1988)]; the KT3 epitope peptide [Martin et al., Science, 255:192-194 (1992)]; an α-tubulin epitope peptide [Skinner et al., J. Biol. Chem., 266:15163-15166 (1991)]; and the T7 gene 10 protein peptide tag [Lutz-Freyermuth et al., Proc. Natl. Acad. Sci. USA, 87:6393-6397 (1990)].

The PRO218 polypeptide of the present invention may also be modified in a way to form a chimeric molecule comprising a PRO218 polypeptide fused to a leucine zipper.

Various leucine zipper polypeptides have been described in the art. See, e.g., Landschulz et al., Science 240:1759 (1988); WO 94/10308; Hoppe et al., FEBS Letters 344:1991 (1994); Maniatis et al., Nature 341:24 (1989). It is believed that use of a leucine zipper fused to a PRO218 polypeptide may be desirable to assist in dimerizing or trimerizing soluble PRO218 polypeptide in solution. Those skilled in the art will appreciate that the leucine zipper may be fused at either the N- or C-terminal end of the PRO218 molecule.

D. Preparation of PRO218

The description below relates primarily to production of PRO218 by culturing cells transformed or transfected with a vector containing PRO218 polypeptide encoding nucleic acid. It is, of course, contemplated that alternative methods, which are well known in the art, may be employed to prepare PRO218 polypeptides. For instance, the PRO218 sequence, or portions thereof, may be produced by direct peptide synthesis using solid-phase techniques [see, e.g., Stewart et al., Solid-Phase Peptide Synthesis, W.H. Freeman Co., San Francisco, CA (1969); Merrifield, J. Am. Chem. Soc., 85:2149-2154 (1963)]. *In vitro* protein synthesis may be performed using manual techniques or by automation. Automated synthesis may be accomplished, for instance, using an Applied Biosystems Peptide Synthesizer (Foster City, CA) using manufacturer's instructions. Various portions of PRO218 polypeptides may be chemically synthesized separately and combined using chemical or enzymatic methods to produce a full-length PRO218 polypeptide.

1. Isolation of DNA Encoding PRO218

DNA encoding a PRO218 polypeptide may be obtained from a cDNA library prepared from tissue believed to possess the PRO218 mRNA and to express it at a detectable level. Accordingly, human PRO218-encoding DNA can be conveniently obtained from a cDNA library prepared from human tissue, such as described in the Examples. The PRO218-encoding gene may also be obtained from a genomic library or by oligonucleotide synthesis.

Libraries can be screened with probes (such as antibodies to a PRO218 polypeptide or oligonucleotides of at least about 20-80 bases) designed to identify the gene of interest or the protein encoded by it. Screening the cDNA or genomic library with the selected probe may be conducted using standard procedures, such as described in Sambrook et al., Molecular Cloning: A Laboratory Manual (New York: Cold Spring Harbor Laboratory Press, 1989). An alternative means to isolate the gene encoding PRO218 is to use PCR methodology [Sambrook et al., supra; Dieffenbach et al., PCR Primer:A Laboratory Manual (Cold Spring

Harbor Laboratory Press, 1995)].

The Examples below describe techniques for screening a cDNA library. The oligonucleotide sequences selected as probes should be of sufficient length and sufficiently unambiguous that false positives are minimized. The oligonucleotide is preferably labeled such that it can be detected upon hybridization to DNA in the library being screened. Methods of labeling are well known in the art, and include the use of radiolabels like ^{32}P-labeled ATP, biotinylation or enzyme labeling. Hybridization conditions, including moderate stringency and high stringency, are provided in Sambrook et al., supra.

Included in this invention are nucleic acids having at least about 70% to about 75% nucleic acid sequence identity with the nucleic acid sequence shown in Figure 1A and 1B (SEQ ID NO:1), with or without the untranslated portions and with or without the region encoding the signal peptide. Ordinarily, in another embodiment, the nucleic acids will have at least about 80% nucleic acid sequence identity, preferably at least about 85% nucleic acid sequence identity, more preferably at least about 90% nucleic acid sequence identity and even more preferably at least about 95% nucleic acid sequence identity with the nucleic acid sequence of Figure 1A and 1B (SEQ ID NO:1), with or without the untranslated portions and with or without the region encoding the signal peptide.

Sequences identified in such library screening methods can be compared and aligned to other known sequences deposited and available in public databases such as GenBank or other private sequence databases. Sequence identity (at either the amino acid or nucleotide level) within defined regions of the molecule or across the full-length sequence can be determined through sequence alignment using computer software programs such as ALIGN, DNAstar, and INHERIT which employ various algorithms to measure homology.

Nucleic acid having protein coding sequence may be obtained by screening selected cDNA or genomic libraries using the deduced amino acid sequence disclosed herein for the first time, and, if necessary, using conventional primer extension procedures as described in Sambrook et al., supra, to detect precursors and processing intermediates of mRNA that may not have been reverse-transcribed into cDNA.

2. Selection and Transformation of Host Cells

Host cells are transfected or transformed with expression or cloning vectors described herein for PRO218 polypeptide production and cultured in conventional nutrient media modified as appropriate for inducing promoters, selecting transformants, or amplifying the genes encoding the desired sequences. The culture conditions, such as media, temperature,

pH and the like, can be selected by the skilled artisan without undue experimentation. In general, principles, protocols, and practical techniques for maximizing the productivity of cell cultures can be found in Mammalian Cell Biotechnology: a Practical Approach, M. Butler, ed. (IRL Press, 1991) and Sambrook et al., supra.

Methods of transfection are known to the ordinarily skilled artisan, for example, $CaPO_4$ and electroporation. Depending on the host cell used, transformation is performed using standard techniques appropriate to such cells. The calcium treatment employing calcium chloride, as described in Sambrook et al., supra, or electroporation is generally used for prokaryotes or other cells that contain substantial cell-wall barriers. Infection with *Agrobacterium tumefaciens* is used for transformation of certain plant cells, as described by Shaw et al., Gene, 23:315 (1983) and WO 89/05859 published 29 June 1989. For mammalian cells without such cell walls, the calcium phosphate precipitation method of Graham and van der Eb, Virology, 52:456-457 (1978) can be employed. General aspects of mammalian cell host system transformations have been described in U.S. Patent No. 4,399,216. Transformations into yeast are typically carried out according to the method of Van Solingen et al., J. Bact., 130:946 (1977) and Hsiao et al., Proc. Natl. Acad. Sci. (USA), 76:3829 (1979). However, other methods for introducing DNA into cells, such as by nuclear microinjection, electroporation, bacterial protoplast fusion with intact cells, or polycations, *e.g.*, polybrene, polyornithine, may also be used. For various techniques for transforming mammalian cells, see Keown et al., Methods in Enzymology, 185:527-537 (1990) and Mansour et al., Nature, 336:348-352 (1988).

Suitable host cells for cloning or expressing the DNA in the vectors herein include prokaryote, yeast, or higher eukaryote cells. Suitable prokaryotes include but are not limited to eubacteria, such as Gram-negative or Gram-positive organisms, for example, Enterobacteriaceae such as *E. coli*. Various *E. coli* strains are publicly available, such as *E. coli* K12 strain MM294 (ATCC 31,446); *E. coli* X1776 (ATCC 31,537); *E. coli* strain W3110 (ATCC 27,325) and K5 772 (ATCC 53,635).

In addition to prokaryotes, eukaryotic microbes such as filamentous fungi or yeast are suitable cloning or expression hosts for PRO218-encoding vectors. *Saccharomyces cerevisiae* is a commonly used lower eukaryotic host microorganism.

Suitable host cells for the expression of glycosylated PRO218 are derived from multicellular organisms. Examples of invertebrate cells include insect cells such as Drosophila S2 and Spodoptera Sf9, as well as plant cells. Examples of useful mammalian host cell lines include Chinese hamster ovary (CHO) and COS cells. More specific examples include

monkey kidney CV1 line transformed by SV40 (COS-7, ATCC CRL 1651); human embryonic kidney line (293 or 293 cells subcloned for growth in suspension culture, Graham et al., J. Gen Virol., 36:59 (1977)); Chinese hamster ovary cells/-DHFR (CHO, Urlaub and Chasin, Proc. Natl. Acad. Sci. USA, 77:4216 (1980)); mouse sertoli cells (TM4, Mather, Biol. Reprod., 23:243-251 (1980)); human lung cells (W138, ATCC CCL 75); human liver cells (Hep G2, HB 8065); and mouse mammary tumor (MMT 060562, ATCC CCL51). The selection of the appropriate host cell is deemed to be within the skill in the art.

3. Selection and Use of a Replicable Vector

The nucleic acid (*e.g.*, cDNA or genomic DNA) encoding the desired PRO218 polypeptide may be inserted into a replicable vector for cloning (amplification of the DNA) or for expression. Various vectors are publicly available. The vector may, for example, be in the form of a plasmid, cosmid, viral particle, or phage. The appropriate nucleic acid sequence may be inserted into the vector by a variety of procedures. In general, DNA is inserted into an appropriate restriction endonuclease site(s) using techniques known in the art. Vector components generally include, but are not limited to, one or more of a signal sequence, an origin of replication, one or more marker genes, an enhancer element, a promoter, and a transcription termination sequence. Construction of suitable vectors containing one or more of these components employs standard ligation techniques which are known to the skilled artisan.

The desired PRO218 polypeptide may be produced recombinantly not only directly, but also as a fusion polypeptide with a heterologous polypeptide, which may be a signal sequence or other polypeptide having a specific cleavage site at the N-terminus of the mature protein or polypeptide. In general, the signal sequence may be a component of the vector, or it may be a part of the PRO218-encoding DNA that is inserted into the vector. The signal sequence may be a prokaryotic signal sequence selected, for example, from the group of the alkaline phosphatase, penicillinase, lpp, or heat-stable enterotoxin II leaders. For yeast secretion the signal sequence may be, *e.g.*, the yeast invertase leader, alpha factor leader (including *Saccharomyces* and *Kluyveromyces* α-factor leaders, the latter described in U.S. Patent No. 5,010,182), or acid phosphatase leader, the *C. albicans* glucoamylase leader (EP 362,179 published 4 April 1990), or the signal described in WO 90/13646 published 15 November 1990. In mammalian cell expression, mammalian signal sequences may be used to direct secretion of the protein, such as signal sequences from secreted polypeptides of the same or related species, as well as viral secretory leaders.

Both expression and cloning vectors contain a nucleic acid sequence that enables the vector to replicate in one or more selected host cells. Such sequences are well known for a variety of bacteria, yeast, and viruses. The origin of replication from the plasmid pBR322 is suitable for most Gram-negative bacteria, the 2μ plasmid origin is suitable for yeast, and various viral origins (SV40, polyoma, adenovirus, VSV or BPV) are useful for cloning vectors in mammalian cells.

Expression and cloning vectors will typically contain a selection gene, also termed a selectable marker. Typical selection genes encode proteins that (a) confer resistance to antibiotics or other toxins, *e.g.*, ampicillin, neomycin, methotrexate, or tetracycline, (b) complement auxotrophic deficiencies, or (c) supply critical nutrients not available from complex media, *e.g.*, the gene encoding D-alanine racemase for *Bacilli*.

An example of suitable selectable markers for mammalian cells are those that enable the identification of cells competent to take up the PRO218-encoding nucleic acid, such as DHFR or thymidine kinase. An appropriate host cell when wild-type DHFR is employed is the CHO cell line deficient in DHFR activity, prepared and propagated as described by Urlaub et al., Proc. Natl. Acad. Sci. USA, 77:4216 (1980). A suitable selection gene for use in yeast is the *trp*1 gene present in the yeast plasmid YRp7 [Stinchcomb et al., Nature, 282:39 (1979); Kingsman et al., Gene, 7:141 (1979); Tschemper et al., Gene, 10:157 (1980)]. The *trp*1 gene provides a selection marker for a mutant strain of yeast lacking the ability to grow in tryptophan, for example, ATCC No. 44076 or PEP4-1 [Jones, Genetics, 85:12 (1977)].

Expression and cloning vectors usually contain a promoter operably linked to the PRO218-encoding nucleic acid sequence to direct mRNA synthesis. Promoters recognized by a variety of potential host cells are well known. Promoters suitable for use with prokaryotic hosts include the β-lactamase and lactose promoter systems [Chang et al., Nature, 275:615 (1978); Goeddel et al., Nature, 281:544 (1979)], alkaline phosphatase, a tryptophan (trp) promoter system [Goeddel, Nucleic Acids Res., 8:4057 (1980); EP 36,776], and hybrid promoters such as the tac promoter [deBoer et al., Proc. Natl. Acad. Sci. USA, 80:21-25 (1983)]. Promoters for use in bacterial systems also will contain a Shine-Dalgarno (S.D.) sequence operably linked to the DNA encoding the PRO218 polypeptide.

Examples of suitable promoting sequences for use with yeast hosts include the promoters for 3-phosphoglycerate kinase [Hitzeman et al., J. Biol. Chem., 255:2073 (1980)] or other glycolytic enzymes [Hess et al., J. Adv. Enzyme Reg., 7:149 (1968); Holland, Biochemistry, 17:4900 (1978)], such as enolase, glyceraldehyde-3-phosphate dehydrogenase, hexokinase, pyruvate decarboxylase, phosphofructokinase, glucose-6-phosphate isomerase, 3-

phosphoglycerate mutase, pyruvate kinase, triosephosphate isomerase, phosphoglucose isomerase, and glucokinase.

Other yeast promoters, which are inducible promoters having the additional advantage of transcription controlled by growth conditions, are the promoter regions for alcohol dehydrogenase 2, isocytochrome C, acid phosphatase, degradative enzymes associated with nitrogen metabolism, metallothionein, glyceraldehyde-3-phosphate dehydrogenase, and enzymes responsible for maltose and galactose utilization. Suitable vectors and promoters for use in yeast expression are further described in EP 73,657.

PRO218 transcription from vectors in mammalian host cells is controlled, for example, by promoters obtained from the genomes of viruses such as polyoma virus, fowlpox virus (UK 2,211,504 published 5 July 1989), adenovirus (such as Adenovirus 2), bovine papilloma virus, avian sarcoma virus, cytomegalovirus, a retrovirus, hepatitis-B virus and Simian Virus 40 (SV40), from heterologous mammalian promoters, *e.g.*, the actin promoter or an immunoglobulin promoter, and from heat-shock promoters, provided such promoters are compatible with the host cell systems.

Transcription of a DNA encoding a PRO218 polypeptide by higher eukaryotes may be increased by inserting an enhancer sequence into the vector. Enhancers are cis-acting elements of DNA, usually about from 10 to 300 bp, that act on a promoter to increase its transcription. Many enhancer sequences are now known from mammalian genes (globin, elastase, albumin, α-fetoprotein, and insulin). Typically, however, one will use an enhancer from a eukaryotic cell virus. Examples include the SV40 enhancer on the late side of the replication origin (bp 100-270), the cytomegalovirus early promoter enhancer, the polyoma enhancer on the late side of the replication origin, and adenovirus enhancers. The enhancer may be spliced into the vector at a position 5' or 3' to the PRO218 coding sequence, but is preferably located at a site 5' from the promoter.

Expression vectors used in eukaryotic host cells (yeast, fungi, insect, plant, animal, human, or nucleated cells from other multicellular organisms) will also contain sequences necessary for the termination of transcription and for stabilizing the mRNA. Such sequences are commonly available from the 5' and, occasionally 3', untranslated regions of eukaryotic or viral DNAs or cDNAs. These regions contain nucleotide segments transcribed as polyadenylated fragments in the untranslated portion of the mRNA encoding PRO218.

Still other methods, vectors, and host cells suitable for adaptation to the synthesis of PRO218 polypeptides in recombinant vertebrate cell culture are described in Gething et al., Nature, 293:620-625 (1981); Mantei et al., Nature, 281:40-46 (1979); EP 117,060; and EP

117,058.

4. Detecting Gene Amplification/Expression

Gene amplification and/or expression may be measured in a sample directly, for example, by conventional Southern blotting, Northern blotting to quantitate the transcription of mRNA [Thomas, Proc. Natl. Acad. Sci. USA, 77:5201-5205 (1980)], dot blotting (DNA analysis), or *in situ* hybridization, using an appropriately labeled probe, based on the sequences provided herein. Alternatively, antibodies may be employed that can recognize specific duplexes, including DNA duplexes, RNA duplexes, and DNA-RNA hybrid duplexes or DNA-protein duplexes. The antibodies in turn may be labeled and the assay may be carried out where the duplex is bound to a surface, so that upon the formation of duplex on the surface, the presence of antibody bound to the duplex can be detected.

Gene expression, alternatively, may be measured by immunological methods, such as immunohistochemical staining of cells or tissue sections and assay of cell culture or body fluids, to quantitate directly the expression of gene product. Antibodies useful for immunohistochemical staining and/or assay of sample fluids may be either monoclonal or polyclonal, and may be prepared in any mammal. Conveniently, the antibodies may be prepared against a native sequence PRO218 polypeptide or against a synthetic peptide based on the DNA sequences provided herein or against exogenous sequence fused to PRO218-encoding DNA and encoding a specific antibody epitope.

5. Purification of Polypeptide

Forms of PRO218 may be recovered from culture medium or from host cell lysates. If membrane-bound, it can be released from the membrane using a suitable detergent solution (*e.g.* Triton-X 100) or by enzymatic cleavage. Cells employed in expression of PRO218 polypeptides can be disrupted by various physical or chemical means, such as freeze-thaw cycling, sonication, mechanical disruption, or cell lysing agents.

It may be desired to purify PRO218 from recombinant cell proteins or polypeptides. The following procedures are exemplary of suitable purification procedures: by fractionation on an ion-exchange column; ethanol precipitation; reverse phase HPLC; chromatography on silica or on a cation-exchange resin such as DEAE; chromatofocusing; SDS-PAGE; ammonium sulfate precipitation; gel filtration using, for example, Sephadex G-75; protein A Sepharose columns to remove contaminants such as IgG; and metal chelating columns to bind epitope-tagged forms of the PRO218 polypeptide. Various methods of protein purification

may be employed and such methods are known in the art and described for example in Deutscher, Methods in Enzymology, 182 (1990); Scopes, Protein Purification:Principles and Practice, Springer-Verlag, New York (1982). The purification step(s) selected will depend, for example, on the nature of the production process used and the particular PRO218 polypeptide produced.

E. Uses for PRO218

Nucleotide sequences (or their complement) encoding PRO218 polypeptides have various applications in the art of molecular biology, including uses as hybridization probes, in chromosome and gene mapping and in the generation of anti-sense RNA and DNA. PRO218-encoding nucleic acid will also be useful for the preparation of PRO218 polypeptides by the recombinant techniques described herein.

The full-length DNA30867-1335 nucleotide sequence (SEQ ID NO:1) or the full-length native sequence PRO218 nucleotide sequence, or portions thereof, may be used as hybridization probes for a cDNA library to isolate the full-length PRO218 gene or to isolate still other genes (for instance, those encoding naturally-occurring variants of PRO218 or PRO218 from other species) which have a desired sequence identity to the PRO218 nucleotide sequence disclosed in Figure 1A and 1B (SEQ ID NO:1). Optionally, the length of the probes will be about 20 to about 50 bases. The hybridization probes may be derived from the UNQ192 (DNA30867-1335) nucleotide sequence of SEQ ID NO:1 as shown in Figure 1A and 1B or from genomic sequences including promoters, enhancer elements and introns of native sequence PRO218-encoding DNA. By way of example, a screening method will comprise isolating the coding region of the PRO218 gene using the known DNA sequence to synthesize a selected probe of about 40 bases. Hybridization probes may be labeled by a variety of labels, including radionucleotides such as ^{32}P or ^{35}S, or enzymatic labels such as alkaline phosphatase coupled to the probe via avidin/biotin coupling systems. Labeled probes having a sequence complementary to that of the PRO218 gene of the present invention can be used to screen libraries of human cDNA, genomic DNA or mRNA to determine which members of such libraries the probe hybridizes to. Hybridization techniques are described in further detail in the Examples below.

The ESTs disclosed in the present application may similarly be employed as probes, using the methods disclosed herein.

The probes may also be employed in PCR techniques to generate a pool of sequences for identification of closely related PRO218 sequences.

Nucleotide sequences encoding a PRO218 polypeptide can also be used to construct hybridization probes for mapping the gene which encodes that PRO218 polypeptide and for the genetic analysis of individuals with genetic disorders. The nucleotide sequences provided herein may be mapped to a chromosome and specific regions of a chromosome using known techniques, such as *in situ* hybridization, linkage analysis against known chromosomal markers, and hybridization screening with libraries.

When the coding sequences for PRO218 encode a protein which binds to another protein (example, where the PRO218 polypeptide functions as a receptor), the PRO218 polypeptide can be used in assays to identify the other proteins or molecules involved in the binding interaction. By such methods, inhibitors of the receptor/ligand binding interaction can be identified. Proteins involved in such binding interactions can also be used to screen for peptide or small molecule inhibitors or agonists of the binding interaction. Also, the receptor PRO218 polypeptide can be used to isolate correlative ligand(s). Screening assays can be designed to find lead compounds that mimic the biological activity of a native PRO218 or a receptor for PRO218. Such screening assays will include assays amenable to high-throughput screening of chemical libraries, making them particularly suitable for identifying small molecule drug candidates. Small molecules contemplated include synthetic organic or inorganic compounds. The assays can be performed in a variety of formats, including protein-protein binding assays, biochemical screening assays, immunoassays and cell based assays, which are well characterized in the art.

Nucleic acids which encode PRO218 polypeptide or any of its modified forms can also be used to generate either transgenic animals or "knock out" animals which, in turn, are useful in the development and screening of therapeutically useful reagents. A transgenic animal (e.g., a mouse or rat) is an animal having cells that contain a transgene, which transgene was introduced into the animal or an ancestor of the animal at a prenatal, e.g., an embryonic stage. A transgene is a DNA which is integrated into the genome of a cell from which a transgenic animal develops. In one embodiment, cDNA encoding PRO218 polypeptide can be used to clone genomic DNA encoding PRO218 in accordance with established techniques and the genomic sequences used to generate transgenic animals that contain cells which express DNA encoding PRO218. Methods for generating transgenic animals, particularly animals such as mice or rats, have become conventional in the art and are described, for example, in U.S. Patent Nos. 4,736,866 and 4,870,009. Typically, particular cells would be targeted for PRO218 transgene incorporation with tissue-specific enhancers. Transgenic animals that include a copy of a transgene encoding PRO218 introduced into the germ line of the animal

at an embryonic stage can be used to examine the effect of increased expression of DNA encoding PRO218. Such animals can be used as tester animals for reagents thought to confer protection from, for example, pathological conditions associated with its overexpression. In accordance with this facet of the invention, an animal is treated with the reagent and a reduced incidence of the pathological condition, compared to untreated animals bearing the transgene, would indicate a potential therapeutic intervention for the pathological condition.

Alternatively, non-human homologues of PRO218 can be used to construct a PRO218 "knock out" animal which has a defective or altered gene encoding PRO218 as a result of homologous recombination between the endogenous gene encoding PRO218 and altered genomic DNA encoding PRO218 introduced into an embryonic cell of the animal. For example, cDNA encoding PRO218 can be used to clone genomic DNA encoding PRO218 in accordance with established techniques. A portion of the genomic DNA encoding PRO218 can be deleted or replaced with another gene, such as a gene encoding a selectable marker which can be used to monitor integration. Typically, several kilobases of unaltered flanking DNA (both at the 5' and 3' ends) are included in the vector [see e.g., Thomas and Capecchi, Cell, 51:503 (1987) for a description of homologous recombination vectors]. The vector is introduced into an embryonic stem cell line (e.g., by electroporation) and cells in which the introduced DNA has homologously recombined with the endogenous DNA are selected [see e.g., Li et al., Cell, 69:915 (1992)]. The selected cells are then injected into a blastocyst of an animal (e.g., a mouse or rat) to form aggregation chimeras [see e.g., Bradley, in *Teratocarcinomas and Embryonic Stem Cells: A Practical Approach*, E. J. Robertson, ed. (IRL, Oxford, 1987), pp. 113-152]. A chimeric embryo can then be implanted into a suitable pseudopregnant female foster animal and the embryo brought to term to create a "knock out" animal. Progeny harboring the homologously recombined DNA in their germ cells can be identified by standard techniques and used to breed animals in which all cells of the animal contain the homologously recombined DNA. Knockout animals can be characterized for instance, for their ability to defend against certain pathological conditions and for their development of pathological conditions due to absence of the PRO218 polypeptide.

F. Anti-PRO218 Antibodies

The present invention further provides anti-PRO218 polypeptide antibodies. Exemplary antibodies include polyclonal, monoclonal, humanized, bispecific, and heteroconjugate antibodies.

1. Polyclonal Antibodies

The anti-PRO218 antibodies of the present invention may comprise polyclonal antibodies. Methods of preparing polyclonal antibodies are known to the skilled artisan. Polyclonal antibodies can be raised in a mammal, for example, by one or more injections of an immunizing agent and, if desired, an adjuvant. Typically, the immunizing agent and/or adjuvant will be injected in the mammal by multiple subcutaneous or intraperitoneal injections. The immunizing agent may include the PRO218 polypeptide or a fusion protein thereof. It may be useful to conjugate the immunizing agent to a protein known to be immunogenic in the mammal being immunized. Examples of such immunogenic proteins include but are not limited to keyhole limpet hemocyanin, serum albumin, bovine thyroglobulin, and soybean trypsin inhibitor. Examples of adjuvants which may be employed include Freund's complete adjuvant and MPL-TDM adjuvant (monophosphoryl Lipid A, synthetic trehalose dicorynomycolate). The immunization protocol may be selected by one skilled in the art without undue experimentation.

2. Monoclonal Antibodies

The anti-PRO218 antibodies may, alternatively, be monoclonal antibodies. Monoclonal antibodies may be prepared using hybridoma methods, such as those described by Kohler and Milstein, Nature, 256:495 (1975). In a hybridoma method, a mouse, hamster, or other appropriate host animal, is typically immunized with an immunizing agent to elicit lymphocytes that produce or are capable of producing antibodies that will specifically bind to the immunizing agent. Alternatively, the lymphocytes may be immunized *in vitro*.

The immunizing agent will typically include the PRO218 polypeptide or a fusion protein thereof. Generally, either peripheral blood lymphocytes ("PBLs") are used if cells of human origin are desired, or spleen cells or lymph node cells are used if non-human mammalian sources are desired. The lymphocytes are then fused with an immortalized cell line using a suitable fusing agent, such as polyethylene glycol, to form a hybridoma cell [Goding, Monoclonal Antibodies: Principles and Practice, Academic Press, (1986) pp. 59-103]. Immortalized cell lines are usually transformed mammalian cells, particularly myeloma cells of rodent, bovine and human origin. Usually, rat or mouse myeloma cell lines are employed. The hybridoma cells may be cultured in a suitable culture medium that preferably contains one or more substances that inhibit the growth or survival of the unfused, immortalized cells. For example, if the parental cells lack the enzyme hypoxanthine guanine phosphoribosyl transferase (HGPRT or HPRT), the culture medium for the hybridomas

typically will include hypoxanthine, aminopterin, and thymidine ("HAT medium"), which substances prevent the growth of HGPRT-deficient cells.

Preferred immortalized cell lines are those that fuse efficiently, support stable high level expression of antibody by the selected antibody-producing cells, and are sensitive to a medium such as HAT medium. More preferred immortalized cell lines are murine myeloma lines, which can be obtained, for instance, from the Salk Institute Cell Distribution Center, San Diego, California and the American Type Culture Collection, Rockville, Maryland. Human myeloma and mouse-human heteromyeloma cell lines also have been described for the production of human monoclonal antibodies [Kozbor, J. Immunol., 133:3001 (1984); Brodeur et al., Monoclonal Antibody Production Techniques and Applications, Marcel Dekker, Inc., New York, (1987) pp. 51-63].

The culture medium in which the hybridoma cells are cultured can then be assayed for the presence of monoclonal antibodies directed against a PRO218 polypeptide. Preferably, the binding specificity of monoclonal antibodies produced by the hybridoma cells is determined by immunoprecipitation or by an *in vitro* binding assay, such as radioimmunoassay (RIA) or enzyme-linked immunoabsorbent assay (ELISA). Such techniques and assays are known in the art. The binding affinity of the monoclonal antibody can, for example, be determined by the Scatchard analysis of Munson and Pollard, Anal. Biochem., 107:220 (1980).

After the desired hybridoma cells are identified, the clones may be subcloned by limiting dilution procedures and grown by standard methods [Goding, supra]. Suitable culture media for this purpose include, for example, Dulbecco's Modified Eagle's Medium and RPMI-1640 medium. Alternatively, the hybridoma cells may be grown *in vivo* as ascites in a mammal.

The monoclonal antibodies secreted by the subclones may be isolated or purified from the culture medium or ascites fluid by conventional immunoglobulin purification procedures such as, for example, protein A-Sepharose, hydroxylapatite chromatography, gel electrophoresis, dialysis, or affinity chromatography.

The monoclonal antibodies may also be made by recombinant DNA methods, such as those described in U.S. Patent No. 4,816,567. DNA encoding the monoclonal antibodies of the invention can be readily isolated and sequenced using conventional procedures (e.g., by using oligonucleotide probes that are capable of binding specifically to genes encoding the heavy and light chains of murine antibodies). The hybridoma cells of the invention serve as a preferred source of such DNA. Once isolated, the DNA may be placed into expression vectors, which are then transfected into host cells such as simian COS cells, Chinese hamster

ovary (CHO) cells, or myeloma cells that do not otherwise produce immunoglobulin protein, to obtain the synthesis of monoclonal antibodies in the recombinant host cells. The DNA also may be modified, for example, by substituting the coding sequence for human heavy and light chain constant domains in place of the homologous murine sequences [U.S. Patent No. 4,816,567; Morrison et al., supra] or by covalently joining to the immunoglobulin coding sequence all or part of the coding sequence for a non-immunoglobulin polypeptide. Such a non-immunoglobulin polypeptide can be substituted for the constant domains of an antibody of the invention, or can be substituted for the variable domains of one antigen-combining site of an antibody of the invention to create a chimeric bivalent antibody.

The antibodies may be monovalent antibodies. Methods for preparing monovalent antibodies are well known in the art. For example, one method involves recombinant expression of immunoglobulin light chain and modified heavy chain. The heavy chain is truncated generally at any point in the Fc region so as to prevent heavy chain crosslinking. Alternatively, the relevant cysteine residues are substituted with another amino acid residue or are deleted so as to prevent crosslinking.

In vitro methods are also suitable for preparing monovalent antibodies. Digestion of antibodies to produce fragments thereof, particularly, Fab fragments, can be accomplished using routine techniques known in the art.

3. Humanized Antibodies

The anti-PRO218 antibodies of the invention may further comprise humanized antibodies or human antibodies. Humanized forms of non-human (e.g., murine) antibodies are chimeric immunoglobulins, immunoglobulin chains or fragments thereof (such as Fv, Fab, Fab', $F(ab')_2$ or other antigen-binding subsequences of antibodies) which contain minimal sequence derived from non-human immunoglobulin. Humanized antibodies include human immunoglobulins (recipient antibody) in which residues from a complementary determining region (CDR) of the recipient are replaced by residues from a CDR of a non-human species (donor antibody) such as mouse, rat or rabbit having the desired specificity, affinity and capacity. In some instances, Fv framework residues of the human immunoglobulin are replaced by corresponding non-human residues. Humanized antibodies may also comprise residues which are found neither in the recipient antibody nor in the imported CDR or framework sequences. In general, the humanized antibody will comprise substantially all of at least one, and typically two, variable domains, in which all or substantially all of the CDR regions correspond to those of a non-human immunoglobulin and all or substantially all of the

FR regions are those of a human immunoglobulin consensus sequence. The humanized antibody optimally also will comprise at least a portion of an immunoglobulin constant region (Fc), typically that of a human immunoglobulin [Jones et al., Nature, 321:522-525 (1986); Riechmann et al., Nature, 332:323-329 (1988); and Presta, Curr. Op. Struct. Biol., 2:593-596 (1992)].

Methods for humanizing non-human antibodies are well known in the art. Generally, a humanized antibody has one or more amino acid residues introduced into it from a source which is non-human. These non-human amino acid residues are often referred to as "import" residues, which are typically taken from an "import" variable domain. Humanization can be essentially performed following the method of Winter and co-workers [Jones et al., Nature, 321:522-525 (1986); Riechmann et al., Nature, 332:323-327 (1988); Verhoeyen et al., Science, 239:1534-1536 (1988)], by substituting rodent CDRs or CDR sequences for the corresponding sequences of a human antibody. Accordingly, such "humanized" antibodies are chimeric antibodies (U.S. Patent No. 4,816,567), wherein substantially less than an intact human variable domain has been substituted by the corresponding sequence from a non-human species. In practice, humanized antibodies are typically human antibodies in which some CDR residues and possibly some FR residues are substituted by residues from analogous sites in rodent antibodies.

Human antibodies can also be produced using various techniques known in the art, including phage display libraries [Hoogenboom and Winter, J. Mol. Biol., 227:381 (1991); Marks et al., J. Mol. Biol., 222:581 (1991)]. The techniques of Cole et al. and Boerner et al. are also available for the preparation of human monoclonal antibodies (Cole et al., Monoclonal Antibodies and Cancer Therapy, Alan R. Liss, p. 77 (1985) and Boerner et al., J. Immunol., 147(1):86-95 (1991)].

4. Bispecific Antibodies

Bispecific antibodies are monoclonal, preferably human or humanized, antibodies that have binding specificities for at least two different antigens. In the present case, one of the binding specificities is for a PRO218 polypeptide, the other one is for any other antigen, and preferably for a cell-surface protein or receptor or receptor subunit.

Methods for making bispecific antibodies are known in the art. Traditionally, the recombinant production of bispecific antibodies is based on the co-expression of two immunoglobulin heavy-chain/light-chain pairs, where the two heavy chains have different specificities [Milstein and Cuello, Nature, 305:537-539 (1983)]. Because of the random

assortment of immunoglobulin heavy and light chains, these hybridomas (quadromas) produce a potential mixture of ten different antibody molecules, of which only one has the correct bispecific structure. The purification of the correct molecule is usually accomplished by affinity chromatography steps. Similar procedures are disclosed in WO 93/08829, published 13 May 1993, and in Traunecker et al., EMBO J., 10:3655-3659 (1991).

Antibody variable domains with the desired binding specificities (antibody-antigen combining sites) can be fused to immunoglobulin constant domain sequences. The fusion preferably is with an immunoglobulin heavy-chain constant domain, comprising at least part of the hinge, CH2, and CH3 regions. It is preferred to have the first heavy-chain constant region (CH1) containing the site necessary for light-chain binding present in at least one of the fusions. DNAs encoding the immunoglobulin heavy-chain fusions and, if desired, the immunoglobulin light chain, are inserted into separate expression vectors, and are co-transfected into a suitable host organism. For further details of generating bispecific antibodies see, for example, Suresh et al., Methods in Enzymology, 121:210 (1986).

5. Heteroconjugate Antibodies

Heteroconjugate antibodies are also within the scope of the present invention. Heteroconjugate antibodies are composed of two covalently joined antibodies. Such antibodies have, for example, been proposed to target immune system cells to unwanted cells [U.S. Patent No. 4,676,980], and for treatment of HIV infection [WO 91/00360; WO 92/200373; EP 03089]. It is contemplated that the antibodies may be prepared *in vitro* using known methods in synthetic protein chemistry, including those involving crosslinking agents. For example, immunotoxins may be constructed using a disulfide exchange reaction or by forming a thioether bond. Examples of suitable reagents for this purpose include iminothiolate and methyl-4-mercaptobutyrimidate and those disclosed, for example, in U.S. Patent No. 4,676,980.

G. Uses for anti-PRO218 Antibodies

The anti-PRO218 antibodies of the present invention have various utilities. For example, anti-PRO218 antibodies may be used in diagnostic assays for PRO218 polypeptides, *e.g.*, detecting expression in specific cells, tissues, or serum. Various diagnostic assay techniques known in the art may be used, such as competitive binding assays, direct or indirect sandwich assays and immunoprecipitation assays conducted in either heterogeneous or homogeneous phases [Zola, Monoclonal Antibodies: A Manual of Techniques, CRC Press,

Inc. (1987) pp. 147-158]. The antibodies used in the diagnostic assays can be labeled with a detectable moiety. The detectable moiety should be capable of producing, either directly or indirectly, a detectable signal. For example, the detectable moiety may be a radioisotope, such as ^{3}H, ^{14}C, ^{32}P, ^{35}S, or ^{125}I, a fluorescent or chemiluminescent compound, such as fluorescein isothiocyanate, rhodamine, or luciferin, or an enzyme, such as alkaline phosphatase, beta-galactosidase or horseradish peroxidase. Any method known in the art for conjugating the antibody to the detectable moiety may be employed, including those methods described by Hunter et al., Nature, 144:945 (1962); David et al., Biochemistry, 13:1014 (1974); Pain et al., J. Immunol. Meth., 40:219 (1981); and Nygren, J. Histochem. and Cytochem., 30:407 (1982).

Anti-PRO218 antibodies also are useful for the affinity purification of PRO218 polypeptides from recombinant cell culture or natural sources. In this process, the antibodies against a PRO218 polypeptide are immobilized on a suitable support, such a Sephadex resin or filter paper, using methods well known in the art. The immobilized antibody then is contacted with a sample containing the PRO218 polypeptide to be purified, and thereafter the support is washed with a suitable solvent that will remove substantially all the material in the sample except the PRO218 polypeptide, which is bound to the immobilized antibody. Finally, the support is washed with another suitable solvent that will release the PRO218 polypeptide from the antibody.

The following examples are offered for illustrative purposes only, and are not intended to limit the scope of the present invention in any way.

All patent and literature references cited in the present specification are hereby incorporated by reference in their entirety.

EXAMPLES

Commercially available reagents referred to in the examples were used according to manufacturer's instructions unless otherwise indicated. The source of those cells identified in the following examples, and throughout the specification, by ATCC accession numbers is the American Type Culture Collection, Rockville, Maryland.

EXAMPLE 1: Isolation of cDNA Clones Encoding Human PRO218

A consensus DNA sequence was assembled manually relative to other EST sequences. This consensus sequence is herein designated DNA-from DNA (also referred to as DNA17411, SEQ ID NO:8) (see Figure 4). The proprietary Genentech EST sequences are

identified in Figure 4 as DNA14472 (SEQ ID NO:3) and DNA15846 (SEQ ID NO:4).

Based on the DNA-from DNA consensus sequence shown in Figure 4, oligonucleotides were synthesized: 1) to identify by PCR a cDNA library that contained the sequence of interest, and 2) for use as probes to isolate a clone of the full-length coding sequence for PRO218. Forward and reverse PCR primers generally range from 20 to 30 nucleotides and are often designed to give a PCR product of about 100-1000 bp in length. The probe sequences are typically 40-55 bp in length. In some cases, additional oligonucleotides are synthesized when the consensus sequence is greater than about 1-1.5kbp. In order to screen several libraries for a full-length clone, DNA from the libraries was screened by PCR amplification, as per Ausubel et al., Current Protocols in Molecular Biology, with the PCR primer pair. A positive library was then used to isolate clones encoding the gene of interest using the probe oligonucleotide and one of the primer pairs.

A pair of PCR primers (forward and reverse) were synthesized:

forward PCR primer 5'-AAGTGGAGCCGGAGCCTTCC-3' (SEQ ID NO:5); and
reverse PCR primer 5'-TCGTTGTTTATGCAGTAGTCGG-3' (SEQ ID NO:6).

Additionally, a synthetic oligonucleotide hybridization probe was constructed from the consensus DNA-from dna sequence which had the following nucleotide sequence:

hybridization probe
5'-ATTGTTTAAAGACTATGAGATACGTCAGTATGTTGTACAGG-3' (SEQ ID NO:7).

In order to screen several libraries for a source of a full-length clone, DNA from the libraries was screened by PCR amplification with the PCR primer pair identified above. A positive library was then used to isolate clones encoding the PRO218 gene using the probe oligonucleotide and one of the PCR primers.

RNA for construction of the cDNA libraries was isolated from human fetal kidney tissue (LIB28). The cDNA libraries used to isolate the cDNA clones were constructed by standard methods using commercially available reagents such as those from Invitrogen, San Diego, CA. The cDNA was primed with oligo dT containing a NotI site, linked with blunt to SalI hemikinased adaptors, cleaved with NotI, sized appropriately by gel electrophoresis, and cloned in a defined orientation into a suitable cloning vector (such as pRKB or pRKD; pRK5B is a precursor of pRK5D that does not contain the SfiI site; see, Holmes et al.,

Science, 253:1278-1280 (1991)) in the unique XhoI and NotI sites.

DNA sequencing of the clones isolated as described above gave the full-length DNA sequence for PRO218 [herein designated as UNQ192 (DNA30867-1335)] (SEQ ID NO:1) and the derived protein sequence for PRO218.

The entire nucleotide sequence of UNQ192 (DNA30867-1335) is shown in Figure 1A and 1B (SEQ ID NO:1). Clone UNQ192 (DNA30867-1335) contains a single open reading frame with an apparent translational initiation site at nucleotide positions 150-152 [Kozak et al., supra] and ending at the stop codon at nucleotide positions 1515-1517 (Figure 1A and 1B). The predicted polypeptide precursor is 455 amino acids long (Figure 2A). The full-length PRO218 protein shown in Figure 2A has an estimated molecular weight of about 52917 and a pI of about 9.5. Clone UNQ192 (DNA30867-1335) has been deposited with the ATCC on 28 April 1998. Regarding the sequence, it is understood that the deposited clone contains the correct sequence, and the sequences provided herein are based on known sequencing techniques.

Analysis of the amino acid sequence of the full-length PRO218 polypeptide suggests that PRO218 may be a novel transmembrane protein.

Still analyzing the amino acid sequence of SEQ ID NO:2, the putative signal peptide is at about amino acids 1 through 23 of SEQ ID NO:2. Transmembrane domains are potentially at about amino acids 37-55, 81-102, 150-168, 288-311, 338-356, 375-398, and 425-444 of SEQ ID NO:2. N-glycosylation sites are at about amino acids 67, 180, and 243 of SEQ ID NO:2. Eukaryotic cobalamin-binding protein is at about amino acids 151-160 of SEQ ID NO:2. The corresponding nucleotides can be routinely determined given the sequences provided herein.

EXAMPLE 2: Use of PRO218-encoding DNA as a hybridization probe

The following method describes use of a nucleotide sequence encoding PRO218 as a hybridization probe.

DNA comprising the coding sequence of full-length PRO218 (as shown in Figure 1A and 1B, SEQ ID NO:1) or a fragment thereof is employed as a probe to screen for homologous DNAs (such as those encoding naturally-occurring variants of PRO218) in human tissue cDNA libraries or human tissue genomic libraries.

Hybridization and washing of filters containing either library DNAs is performed under the following high stringency conditions. Hybridization of radiolabeled PRO218 polypeptide-derived probe to the filters is performed in a solution of 50% formamide, 5 x SSC, 0.1% SDS,

0.1% sodium pyrophosphate, 50 mM sodium phosphate, pH 6.8, 2 x Denhardt's solution, and 10% dextran sulfate at 42°C for 20 hours. Washing of the filters is performed in an aqueous solution of 0.1 x SSC and 0.1% SDS at 42°C.

DNAs having a desired sequence identity with the DNA encoding full-length native sequence PRO218 polypeptide can then be identified using standard techniques known in the art.

EXAMPLE 3: Expression of PRO218 Polypeptides in *E. coli*

This example illustrates the preparation of unglycosylated forms of PRO218 polypeptides by recombinant expression in *E. coli*.

The DNA sequence encoding the full-length PRO218 or a fragment or variant thereof is initially amplified using selected PCR primers. The primers should contain restriction enzyme sites which correspond to the restriction enzyme sites on the selected expression vector. A variety of expression vectors may be employed. An example of a suitable vector is pBR322 (derived from *E. coli*; see Bolivar et al., Gene, 2:95 (1977)) which contains genes for ampicillin and tetracycline resistance. The vector is digested with restriction enzyme and dephosphorylated. The PCR amplified sequences are then ligated into the vector. The vector will preferably include sequences which encode for an antibiotic resistance gene, a trp promoter, a polyhis leader (including the first six STII codons, polyhis sequence, and enterokinase cleavage site), the PRO218 coding region, lambda transcriptional terminator, and an argU gene.

The ligation mixture is then used to transform a selected *E. coli* strain using the methods described in Sambrook et al., supra. Transformants are identified by their ability to grow on LB plates and antibiotic resistant colonies are then selected. Plasmid DNA can be isolated and confirmed by restriction analysis and DNA sequencing.

Selected clones can be grown overnight in liquid culture medium such as LB broth supplemented with antibiotics. The overnight culture may subsequently be used to inoculate a larger scale culture. The cells are then grown to a desired optical density, during which the expression promoter is turned on.

After culturing the cells for several more hours, the cells can be harvested by centrifugation. The cell pellet obtained by the centrifugation can be solubilized using various agents known in the art, and the solubilized PRO218 polypeptide can then be purified using a metal chelating column under conditions that allow tight binding of the polypeptide.

EXAMPLE 4: Expression of PRO218 Polypeptides in Mammalian Cells

This example illustrates preparation of glycosylated forms of PRO218 polypeptides by recombinant expression in mammalian cells.

The vector, pRK5 (see EP 307,247, published March 15, 1989), is employed as the expression vector. Optionally, the PRO218-encoding DNA is ligated into pRK5 with selected restriction enzymes to allow insertion of the PRO218-encoding DNA using ligation methods such as described in Sambrook et al., supra. The resulting vector is called pRK5-PRO218.

In one embodiment, the selected host cells may be 293 cells. Human 293 cells (ATCC CCL 1573) are grown to confluence in tissue culture plates in medium such as DMEM supplemented with fetal calf serum and optionally, nutrient components and/or antibiotics. About 10 μg pRK5-PRO218 DNA is mixed with about 1 μg DNA encoding the VA RNA gene [Thimmappaya et al., Cell, 31:543 (1982)] and dissolved in 500 μl of 1 mM Tris-HCl, 0.1 mM EDTA, 0.227 M $CaCl_2$. To this mixture is added, dropwise, 500 μl of 50 mM HEPES (pH 7.35), 280 mM NaCl, 1.5 mM $NaPO_4$, and a precipitate is allowed to form for 10 minutes at 25°C. The precipitate is suspended and added to the 293 cells and allowed to settle for about four hours at 37°C. The culture medium is aspirated off and 2 ml of 20% glycerol in PBS is added for 30 seconds. The 293 cells are then washed with serum free medium, fresh medium is added and the cells are incubated for about 5 days.

Approximately 24 hours after the transfections, the culture medium is removed and replaced with culture medium (alone) or culture medium containing 200 μCi/ml ^{35}S-cysteine and 200 μCi/ml ^{35}S-methionine. After a 12 hour incubation, the conditioned medium is collected, concentrated on a spin filter, and loaded onto a 15% SDS gel. The processed gel may be dried and exposed to film for a selected period of time to reveal the presence of PRO218 polypeptide. The cultures containing transfected cells may undergo further incubation (in serum free medium) and the medium is tested in selected bioassays.

In an alternative technique, PRO218-encoding DNA may be introduced into 293 cells transiently using the dextran sulfate method described by Somparyrac et al., Proc. Natl. Acad. Sci., 12:7575 (1981). 293 cells are grown to maximal density in a spinner flask and 700 μg pRK5-PRO218 DNA is added. The cells are first concentrated from the spinner flask by centrifugation and washed with PBS. The DNA-dextran precipitate is incubated on the cell pellet for four hours. The cells are treated with 20% glycerol for 90 seconds, washed with tissue culture medium, and re-introduced into the spinner flask containing tissue culture medium, 5 μg/ml bovine insulin and 0.1 μg/ml bovine transferrin. After about four days, the conditioned media is centrifuged and filtered to remove cells and debris. The sample

containing expressed PRO218 polypeptide can then be concentrated and purified by any selected method, such as dialysis and/or column chromatography.

In another embodiment, PRO218 polypeptide can be expressed in CHO cells. The pRK5-PRO218 vector can be transfected into CHO cells using known reagents such as $CaPO_4$ or DEAE-dextran. As described above, the cell cultures can be incubated, and the medium replaced with culture medium (alone) or medium containing a radiolabel such as ^{35}S-methionine. After determining the presence of PRO218 polypeptide, the culture medium may be replaced with serum free medium. Preferably, the cultures are incubated for about 6 days, and then the conditioned medium is harvested. The medium containing the expressed PRO218 polypeptide can then be concentrated and purified by any selected method.

Epitope-tagged PRO218 polypeptide may also be expressed in host CHO cells. The PRO218-encoding DNA may be subcloned out of the pRK5 vector. The subclone insert can undergo PCR to fuse in frame with a selected epitope tag such as a poly-his tag into a Baculovirus expression vector. The poly-his tagged PRO218-encoding DNA insert can then be subcloned into a SV40 driven vector containing a selection marker such as DHFR for selection of stable clones. Finally, the CHO cells can be transfected (as described above) with the SV40 driven vector. Labeling may be performed, as described above, to verify expression. The culture medium containing the expressed poly-His tagged PRO218 polypeptide can then be concentrated and purified by any selected method, such as by Ni^{2+}-chelate affinity chromatography.

EXAMPLE 5: Expression of a PRO218 Polypeptide in Yeast

The following method describes recombinant expression of PRO218 polypeptides in yeast.

First, yeast expression vectors are constructed for intracellular production or secretion of PRO218 polypeptide from the ADH2/GAPDH promoter. DNA encoding the PRO218 polypeptide of interest, a selected signal peptide and the promoter is inserted into suitable restriction enzyme sites in the selected plasmid to direct intracellular expression of the PRO218 polypeptide. For secretion, DNA encoding the PRO218 polypeptide can be cloned into the selected plasmid, together with DNA encoding the ADH2/GAPDH promoter, the yeast alpha-factor secretory signal/leader sequence, and linker sequences (if needed) for expression of the PRO218 polypeptide.

Yeast cells, such as yeast strain AB110, can then be transformed with the expression plasmids described above and cultured in selected fermentation media. The transformed yeast

supernatants can be analyzed by precipitation with 10% trichloroacetic acid and separation by SDS-PAGE, followed by staining of the gels with Coomassie Blue stain.

Recombinant PRO218 polypeptide can subsequently be isolated and purified by removing the yeast cells from the fermentation medium by centrifugation and then concentrating the medium using selected cartridge filters. The concentrate containing the PRO218 polypeptide may further be purified using selected column chromatography resins.

EXAMPLE 6: Expression of PRO218 Polypeptides in Baculovirus-Infected Insect Cells

The following method describes recombinant expression of PRO218 polypeptides in Baculovirus-infected insect cells.

The PRO218-encoding DNA is fused upstream of an epitope tag contained within a baculovirus expression vector. Such epitope tags include poly-his tags and immunoglobulin tags (like Fc regions of IgG). A variety of plasmids may be employed, including plasmids derived from commercially available plasmids such as pVL1393 (Novagen). Briefly, the PRO218-encoding DNA or the desired portion of the PRO218-encoding DNA (such as the sequence encoding the extracellular domain of a transmembrane protein) is amplified by PCR with primers complementary to the 5' and 3' regions. The 5' primer may incorporate flanking (selected) restriction enzyme sites. The product is then digested with those selected restriction enzymes and subcloned into the expression vector.

Recombinant baculovirus is generated by co-transfecting the above plasmid and BaculoGold™ virus DNA (Pharmingen) into *Spodoptera frugiperda* ("Sf9") cells (ATCC CRL 1711) using lipofectin (commercially available from GIBCO-BRL). After 4 to 5 days of incubation at 28°C, the released viruses are harvested and used for further amplifications. Viral infection and protein expression is performed as described by O'Reilley et al., Baculovirus expression vectors: A laboratory Manual, Oxford:Oxford University Press (1994).

Expressed poly-his tagged PRO218 polypeptide can then be purified, for example, by Ni^{2+}-chelate affinity chromatography as follows. Extracts are prepared from recombinant virus-infected Sf9 cells as described by Rupert et al., Nature, 362:175-179 (1993). Briefly, Sf9 cells are washed, resuspended in sonication buffer (25 mL Hepes, pH 7.9; 12.5 mM $MgCl_2$; 0.1 mM EDTA; 10% Glycerol; 0.1% NP-40; 0.4 M KCl), and sonicated twice for 20 seconds on ice. The sonicates are cleared by centrifugation, and the supernatant is diluted 50-fold in loading buffer (50 mM phosphate, 300 mM NaCl, 10% Glycerol, pH 7.8) and filtered through a 0.45 μm filter. A Ni^{2+}-NTA agarose column (commercially available from Qiagen) is prepared with a bed volume of 5 mL, washed with 25 mL of water and equilibrated with

25 mL of loading buffer. The filtered cell extract is loaded onto the column at 0.5 mL per minute. The column is washed to baseline A_{280} with loading buffer, at which point fraction collection is started. Next, the column is washed with a secondary wash buffer (50 mM phosphate; 300 mM NaCl, 10% Glycerol, pH 6.0), which elutes nonspecifically bound protein. After reaching A_{280} baseline again, the column is developed with a 0 to 500 mM Imidazole gradient in the secondary wash buffer. One mL fractions are collected and analyzed by SDS-PAGE and silver staining or western blot with Ni^{2+}-NTA-conjugated to alkaline phosphatase (Qiagen). Fractions containing the eluted His_{10}-tagged PRO218 polypeptide are pooled and dialyzed against loading buffer.

Alternatively, purification of the IgG tagged (or Fc tagged) PRO218 polypeptide can be performed using known chromatography techniques, including for instance, Protein A or protein G column chromatography.

EXAMPLE 7: Preparation of Antibodies that Bind PRO218 Polypeptides

This example illustrates the preparation of monoclonal antibodies which can specifically bind to PRO218 polypeptides.

Techniques for producing the monoclonal antibodies are known in the art and are described, for instance, in Goding, supra. Immunogens that may be employed include purified PRO218 polypeptide, fusion proteins containing a PRO218 polypeptide, and cells expressing recombinant PRO218 polypeptide on the cell surface. Selection of the immunogen can be made by the skilled artisan without undue experimentation.

Mice, such as Balb/c, are immunized with the PRO218 immunogen emulsified in complete Freund's adjuvant and injected subcutaneously or intraperitoneally in an amount from 1-100 micrograms. Alternatively, the immunogen is emulsified in MPL-TDM adjuvant (Ribi Immunochemical Research, Hamilton, MT) and injected into the animal's hind foot pads. The immunized mice are then boosted 10 to 12 days later with additional immunogen emulsified in the selected adjuvant. Thereafter, for several weeks, the mice may also be boosted with additional immunization injections. Serum samples may be periodically obtained from the mice by retro-orbital bleeding for testing in ELISA assays to detect anti-PRO218 polypeptide antibodies.

After a suitable antibody titer has been detected, the animals "positive" for antibodies can be injected with a final intravenous injection of PRO218 polypeptide. Three to four days later, the mice are sacrificed and the spleen cells are harvested. The spleen cells are then fused (using 35% polyethylene glycol) to a selected murine myeloma cell line such as

P3X63AgU.1, available from ATCC, No. CRL 1597. The fusions generate hybridoma cells which can then be plated in 96 well tissue culture plates containing HAT (hypoxanthine, aminopterin, and thymidine) medium to inhibit proliferation of non-fused cells, myeloma hybrids, and spleen cell hybrids.

The hybridoma cells will be screened in an ELISA for reactivity against PRO218 polypeptide. Determination of "positive" hybridoma cells secreting the desired monoclonal antibodies against a PRO218 polypeptide is within the skill in the art.

The positive hybridoma cells can be injected intraperitoneally into syngeneic Balb/c mice to produce ascites containing the anti-PRO218 polypeptide monoclonal antibodies. Alternatively, the hybridoma cells can be grown in tissue culture flasks or roller bottles. Purification of the monoclonal antibodies produced in the ascites can be accomplished using ammonium sulfate precipitation, followed by gel exclusion chromatography. Alternatively, affinity chromatography based upon binding of antibody to protein A or protein G can be employed.

EXAMPLE 8: Identification of SEQ ID NOS:3 and 4 Using Amylase Screening

1. Preparation of oligo dT primed cDNA library

mRNA was isolated from human fetal kidney tissue using reagents and protocols from Invitrogen, San Diego, CA (Fast Track 2). This RNA was used to generate an oligo dT primed cDNA library in the vector pRK5D using reagents and protocols from Life Technologies, Gaithersburg, MD (Super Script Plasmid System). In this procedure, the double stranded cDNA was sized to greater than 1000 bp and the SalI/NotI linkered cDNA was cloned into XhoI/NotI cleaved vector. pRK5D is a cloning vector that has an sp6 transcription initiation site followed by an SfiI restriction enzyme site preceding the XhoI/NotI cDNA cloning sites.

2. Preparation of random primed cDNA library

A secondary cDNA library was generated in order to preferentially represent the 5' ends of the primary cDNA clones. Sp6 RNA was generated from the primary library (described above), and this RNA was used to generate a random primed cDNA library in the vector pSST-AMY.0 using reagents and protocols from Life Technologies (Super Script Plasmid System, referenced above). In this procedure the double stranded cDNA was sized to 500-1000 bp, linkered with blunt to NotI adaptors, cleaved with SfiI, and cloned into SfiI/NotI cleaved vector. pSST-AMY.0 is a cloning vector that has a yeast alcohol

dehydrogenase promoter preceding the cDNA cloning sites and the mouse amylase sequence (the mature sequence without the secretion signal) followed by the yeast alcohol dehydrogenase terminator, after the cloning sites. Thus, cDNAs cloned into this vector that are fused in frame with amylase sequence will lead to the secretion of amylase from appropriately transfected yeast colonies.

3. Transformation and Detection

DNA from the library described in paragraph 2 above was chilled on ice to which was added electrocompetent DH10B bacteria (Life Technologies, 20 ml). The bacteria and vector mixture was then electroporated as recommended by the manufacturer. Subsequently, SOC media (Life Technologies, 1 ml) was added and the mixture was incubated at 37°C for 30 minutes. The transformants were then plated onto 20 standard 150 mm LB plates containing ampicillin and incubated for 16 hours (37°C). Positive colonies were scraped off the plates and the DNA was isolated from the bacterial pellet using standard protocols, e.g. CsCl-gradient. The purified DNA was then carried on to the yeast protocols below.

The yeast methods were divided into three categories: (1) Transformation of yeast with the plasmid/cDNA combined vector; (2) Detection and isolation of yeast clones secreting amylase; and (3) PCR amplification of the insert directly from the yeast colony and purification of the DNA for sequencing and further analysis.

The yeast strain used was HD56-5A (ATCC-90785). This strain has the following genotype: MAT alpha, ura3-52, leu2-3, leu2-112, his3-11, his3-15, MAL^+, SUC^+, GAL^+.

Transformation was performed based on the protocol outlined by Gietz et al., Nucl. Acid. Res., 20:1425 (1992). Transformed cells were then inoculated from agar into YEPD complex media broth (100 ml) and grown overnight at 30°C. The YEPD broth was prepared as described in Kaiser et al., Methods in Yeast Genetics, Cold Spring Harbor Press, Cold Spring Harbor, NY, p. 207 (1994). The overnight culture was then diluted to about 2×10^6 cells/ml (approx. $OD_{600}=0.1$) into fresh YEPD broth (500 ml) and regrown to 1×10^7 cells/ml (approx. $OD_{600}=0.4$-0.5).

The cells were then harvested and prepared for transformation by transfer into GS3 rotor bottles in a Sorval GS3 rotor at 5,000 rpm for 5 minutes, the supernatant discarded, and then resuspended into sterile water, and centrifuged again in 50 ml falcon tubes at 3,500 rpm in a Beckman GS-6KR centrifuge. The supernatant was discarded and the cells were subsequently washed with LiAc/TE (10 ml, 10 mM Tris-HCl, 1 mM EDTA pH 7.5, 100 mM Li_2OOCCH_3), and resuspended into LiAc/TE (2.5 ml).

Transformation took place by mixing the prepared cells (100 μl) with freshly denatured single stranded salmon testes DNA (Lofstrand Labs, Gaithersburg, MD) and transforming DNA (1 μg, vol. < 10 μl) in microfuge tubes. The mixture was mixed briefly by vortexing, then 40% PEG/TE (600 μl, 40% polyethylene glycol-4000, 10 mM Tris-HCl, 1 mM EDTA, 100 mM Li_2OOCCH_3, pH 7.5) was added. This mixture was gently mixed and incubated at 30°C while agitating for 30 minutes. The cells were then heat shocked at 42°C for 15 minutes, and the reaction vessel centrifuged in a microfuge at 12,000 rpm for 5-10 seconds, decanted and resuspended into TE (500 μl, 10 mM Tris-HCl, 1 mM EDTA pH 7.5) followed by recentrifugation. The cells were then diluted into TE (1 ml) and aliquots (200 μl) were spread onto the selective media previously prepared in 150 mm growth plates (VWR).

Alternatively, instead of multiple small reactions, the transformation was performed using a single, large scale reaction, wherein reagent amounts were scaled up accordingly.

The selective media used was a synthetic complete dextrose agar lacking uracil (SCD-Ura) prepared as described in Kaiser et al., Methods in Yeast Genetics, Cold Spring Harbor Press, Cold Spring Harbor, NY, p. 208-210 (1994). Transformants were grown at 30°C for 2-3 days.

The detection of colonies secreting amylase was performed by including red starch in the selective growth media. Starch was coupled to the red dye (Reactive Red-120, Sigma) as per the procedure described by Biely et al., Anal. Biochem., 172:176-179 (1988). The coupled starch was incorporated into the SCD-Ura agar plates at a final concentration of 0.15% (w/v), and was buffered with potassium phosphate to a pH of 7.0 (50-100 mM final concentration).

The positive colonies were picked and streaked across fresh selective media (onto 150 mm plates) in order to obtain well isolated and identifiable single colonies. Well isolated single colonies positive for amylase secretion were detected by direct incorporation of red starch into buffered SCD-Ura agar. Positive colonies were determined by their ability to break down starch resulting in a clear halo around the positive colony visualized directly.

4. Isolation of DNA by PCR Amplification

When a positive colony was isolated, a portion of it was picked by a toothpick and diluted into sterile water (30 μl) in a 96 well plate. At this time, the positive colonies were either frozen and stored for subsequent analysis or immediately amplified.

An aliquot of the reactions (5 μl) were examined by agarose gel electrophoresis in a 1% agarose gel using a Tris-Borate-EDTA (TBE) buffering system as described by Sambrook

et al., supra. Clones resulting in a single strong PCR product larger than 400 bp were further analyzed by DNA sequencing after purification with a 96 Qiaquick PCR clean-up column (Qiagen Inc., Chatsworth, CA). Sequencing identified novel ESTs having SEQ ID NO:3 or 4.

Deposit of Material

The following materials have been deposited with the American Type Culture Collection, 12301 Parklawn Drive, Rockville, MD, USA (ATCC):

Material	ATCC Dep. No.	Deposit Date
DNA30867-1335		28 April 1998

This deposit was made under the provisions of the Budapest Treaty on the International Recognition of the Deposit of Microorganisms for the Purpose of Patent Procedure and the Regulations thereunder (Budapest Treaty). This assures maintenance of a viable culture of the deposit for 30 years from the date of deposit. The deposit will be made available by ATCC under the terms of the Budapest Treaty, and subject to an agreement between Genentech, Inc. and ATCC, which assures permanent and unrestricted availability of the progeny of the culture of the deposit to the public upon issuance of the pertinent U.S. patent or upon laying open to the public of any U.S. or foreign patent application, whichever comes first, and assures availability of the progeny to one determined by the U.S. Commissioner of Patents and Trademarks to be entitled thereto according to 35 USC §122 and the Commissioner's rules pursuant thereto (including 37 CFR §1.14 with particular reference to 886 OG 638).

The assignee of the present application has agreed that if a culture of the materials on deposit should die or be lost or destroyed when cultivated under suitable conditions, the materials will be promptly replaced on notification with another of the same. Availability of the deposited material is not to be construed as a license to practice the invention in contravention of the rights granted under the authority of any government in accordance with its patent laws.

The foregoing written specification is considered to be sufficient to enable one skilled in the art to practice the invention. The present invention is not to be limited in scope by the construct deposited, since the deposited embodiment is intended as a single illustration of certain aspects of the invention and any constructs that are functionally equivalent are within the scope of this invention. The deposit of material herein does not constitute an admission that the written description herein contained is inadequate to enable the practice of any aspect

of the invention, including the best mode thereof, nor is it to be construed as limiting the scope of the claims to the specific illustrations that it represents. Indeed, various modifications of the invention in addition to those shown and described herein will become apparent to those skilled in the art from the foregoing description and fall within the scope of the appended claims.

WHAT IS CLAIMED IS:

1. Isolated nucleic acid comprising DNA having at least an 80% sequence identity to (a) a DNA molecule encoding a PRO218 polypeptide comprising the sequence of amino acid residues 1 or 24 through 455 of Figure 2A (SEQ ID NO:2), or (b) the complement of the DNA molecule of (a).

2. The nucleic acid of Claim 1, wherein said DNA comprises nucleotides 150 or 219 through 1514 of the nucleotide sequence of SEQ ID NO:1.

3. The nucleic acid of Claim 1, wherein said DNA comprises the nucleotide sequence of SEQ ID NO:1.

4. Isolated nucleic acid comprising DNA having at least an 80% sequence identity to (a) a DNA molecule encoding the same mature polypeptide encoded by the human protein cDNA in ATCC Deposit of DNA30867-1335, or (b) the complement of the DNA molecule of (a).

5. The nucleic acid of Claim 4 which comprises a DNA molecule encoding the same mature polypeptide encoded by the human protein cDNA in said ATCC Deposit.

6. A vector comprising the nucleic acid of any one of Claims 1 through 6.

7. The vector of Claim 6 operably linked to control sequences recognized by a host cell transformed with the vector.

8. A host cell comprising the vector of Claim 6.

9. The host cell of Claim 8, wherein said cell is a CHO cell.

10. The host cell of Claim 8, wherein said cell is an *E. coli*.

11. The host cell of Claim 8, wherein said cell is a yeast cell.

12. A process for producing a PRO218 polypeptide comprising culturing the host

cell of Claim 8 under conditions suitable for expression of said PRO218 polypeptide and recovering said PRO218 polypeptide from the cell culture.

13. Isolated native sequence PRO218 polypeptide comprising amino acid residues 1 or 24 through 455 of Figure 2 (SEQ ID NO:2).

14. Isolated PRO218 polypeptide encoded by the cDNA insert of the vector deposited as DNA30867-1335 with the ATCC.

15. A chimeric molecule comprising a PRO218 polypeptide fused to a heterologous amino acid sequence.

16. The chimeric molecule of Claim 15, wherein said heterologous amino acid sequence is an epitope tag sequence.

17. The chimeric molecule of Claim 15, wherein said heterologous amino acid sequence is a Fc region of an immunoglobulin.

18. An antibody which specifically binds to a PRO218 polypeptide.

19. The antibody of Claim 18, wherein said antibody is a monoclonal antibody.

20. An isolated nucleic acid comprising the nucleotide sequence of SEQ ID NO:3 or its complement.

21. An isolated nucleic acid comprising the nucleotide sequence of SEQ ID NO:4 or its complement.

Abstract of the Disclosure

The present invention is directed to novel polypeptides having homology to membrane regulator proteins and to nucleic acid molecules encoding those polypeptides. Also provided herein are vectors and host cells comprising those nucleic acid sequences, chimeric polypeptide molecules comprising the polypeptides of the present invention fused to heterologous polypeptide sequences, antibodies which bind to the polypeptides of the present invention and to methods for producing the polypeptides of the present invention.

> DNAJ0867 [Min]
> 0 Sites [No Sites]
> length: 2037 bp (circular)

SEQ. ID NO: 1

```
  1-1 CGGACGCGTG GGCGGACGCG TGGGGGAGAG CCGCAGTCCC GGCTGCAGCA CCTGGGAGAA GGCAGACCGT GTGAGGGGGC CTGTGGCCCC AGCGTGCTGT
      GCCTGCGCAC CCGCCTGCGC ACCCCCTCTC GGCGTCAGGG CCGACGTCGT GGACCCTCTT CCGTCTGGCA CACTCCCCCG GACACCGGGG TCGCACGACA

 101 GGCCTCGGGG AGTGGGAAGT GGAGGCAGGA GCCTTCCTTA CACTTCGCCA TGAGTTTCCT CATCGACTCC AGCATCATGA TTACCTCCCA GATACTATTT
     CCGGAGCCCC TCACCCTTCA CCTCCGTCCT CGGAAGGAAT GTGAAGCGGT ACTCAAAGGA GTAGCTGAGG TCGTAGTACT AATGGAGGGT CTATGATAAA
   1                                                    M  S  F  L  I  D  S  S  I  M  I  T  S  Q  I  L  F
                                                        ^Start

 201 TTTGGATTTG GGTGGCTTTT CTTCATGCGC CAATTGTTTA AAGACTATGA GATACGTCAG TATGTTGTAC AGGTGATCTT CTCCGTGACG TTTGCATTTT
     AAACCTAAAC CCACCGAAAA GAAGTACGCG GTTAACAAAT TTCTGATACT CTATGCAGTC ATACAACATG TCCACTAGAA GAGGCACTGC AAACGTAAAA
  18 F  G  F  G   W  L  F   F  M  R   Q  L  F  K   D  Y  E   I  R  Q   Y  V  V  Q   V  I  F   S  V  T   F  A  F  S

 301 CTTGCACCAT GTTTGAGCTC ATCATCTTTG AAATCTTAGG AGTATTGAAT AGCAGCTCCC GTTATTTTCA CTGGAAAATG AACCTGTGTG TAATTCTGCT
     GAACGTGGTA CAAACTCGAG TAGTAGAAAC TTTAGAATCC TCATAACTTA TCGTCGAGGG CAATAAAAGT GACCTTTTAC TTGGACACAC ATTAAGACGA
  52   C  T  M   F  E  L   I  I  F  E   I  L  G   V  L  N   S  S  S  R   Y  F  H   W  K  M   N  L  C  V   I  L  L

 401 GATCCTGGTT TTCATGGTGC CTTTTTACAT TGGCTATTTT ATTGTGAGCA ATATCCGACT ACTGCATAAA CAACGACTGC TTTTTTCCTG TCTCTTATGG
     CTAGGACCAA AAGTACCACG GAAAAATGTA ACCGATAAAA TAACACTCGT TATAGGCTGA TGACGTATTT GTTGCTGACG AAAAAAGGAC AGAGAATACC
  85  I  L  V   F  M  V  P   F  Y  I   G  Y  F   I  V  S  N   I  R  L   L  H  K   Q  R  L  L   P  S  C   L  L  W

 501 CTGACCTTTA TGTATTTCTT CTGGAAACTA GGAGATCCCT TTCCCATTCT CAGCCCAAAA CATGGGATCT TATCCATAGA ACAGCTCATC AGCCGGGTTG
     GACTGGAAAT ACATAAAGAA GACCTTTGAT CCTCTAGGGA AAGGGTAAGA GTCGGGTTTT GTACCCTAGA ATAGGTATCT TGTCGAGTAG TCGGCCCAAC
 118 L  T  F  M   Y  F  F   W  K  L   G  D  P  F   P  I  L   S  P  K   H  G  I  L   S  I  E   Q  L  I   S  R  V  G

 601 GTGTGATTGG AGTGACTCTC ATGGCTCTTC TTTCTGGATT TGGTGCTGTC AACTGCCCAT ACACTTACAT GTCTTACTTC CTCAGGAATG TGACTGACAC
     CACACTAACC TCACTGAGAG TACCGAGAAG AAAGACCTAA ACCACGACAG TTGACGGGTA TGTGAATGTA CAGAATGAAG GAGTCCTTAC ACTGACTGTG
 152   V  I  G   V  T  L   M  A  L  L   S  G  P   G  A  V   N  C  P  Y   T  Y  M   S  Y  F   L  R  N  V   T  D  T

 701 GGATATTCTA GCCCTGGAAC GGCGACTGCT GCAAACCATG GATATGATCA TAAGCAAAAA GAAAAGGATG GCAATGGCAC GGAGAACAAT GTTCCAGAAG
     CCTATAAGAT CGGGACCTTG CCGCTGACGA CGTTTGGTAC CTATACTAGT ATTCGTTTTT CTTTTCCTAC CGTTACCGTG CCTCTTGTTA CAAGGTCTTC
 185  D  I  L   A  L  E  R   R  L  L   Q  T  M   D  M  I  I   S  K  K   K  R  M   A  M  A  R   R  T  M   F  Q  K

 801 GGGGAAGTGC ATAACAAACC ATCAGGTTTC TGGGGAATGA TAAAAAGTGT TACCACTTCA GCATCAGGAA GTGAAAATCT TACTCTTATT CAACAGGAAG
     CCCCTTCACG TATTGTTTGG TAGTCCAAAG ACCCCTTACT ATTTTTCACA ATGGTGAAGT CGTAGTCCTT CACTTTTAGA ATGAGAATAA GTTGTCCTTC
 218 G  E  V  H   N  K  P   S  G  F   W  G  M  I   K  S  V   T  T  S   A  S  G  S   E  N  L   T  L  I   Q  Q  E  V
```

FIGURE 1A

```
 901 TGGATGCTTT GGAAGAATTA AGCAGGCAGC TTTTTCTGGA AACAGCTGAT CTATATGCTA CCAAGGAGAG AATAGAATAC TCCAAAACCT TCAAGGGGAA
     ACCTACGAAA CCTTCTTAAT TCGTCCGTCG AAAAAGACCT TTGTCGACTA GATATACGAT GGTTCCTCTC TTATCTTATG AGGTTTTGGA AGTTCCCCTT
 252   D  A  L   E  E  L   S  R  Q  L   F  L  E   T  A  D   L  Y  A  T   K  E  R   I  E  Y   S  K  T  F   K  G  K

1001 ATATTTTAAT TTTCTTGGTT ACTTTTTCTC TATTTACTGT GTTTGGAAAA TTTTCATGGC TACCATCAAT ATTGTTTTTG ATCGAGTTGG GAAAACGGAT
     TATAAAATTA AAAGAACCAA TGAAAAAGAG ATAAATGACA CAAACCTTTT AAAAGTACCG ATGGTAGTTA TAACAAAAAC TAGCTCAACC CTTTTGCCTA
 285 Y  F  N   F  L  G  Y   F  F  S   I  Y  C   V  W  K  I   F  M  A   T  I  N   I  V  F  D   R  V  G   K  T  D

1101 CCTGTCACAA GAGGCATTGA GATCACTGTG AATTATCTGG GAATCCAATT TGATGTGAAG TTTTGGTCCC AACACATTTC CTTCATTCTT GTTGGAATAA
     GGACAGTGTT CTCCGTAACT CTAGTGACAC TTAATAGACC CTTAGGTTAA ACTACACTTC AAAACCAGGG TTGTGTAAAG GAAGTAAGAA CAACCTTATT
 318 P  V  T  R   G  I  E   I  T  V   N  Y  L  G   I  Q  F   D  V  K   F  W  S  Q   H  I  S   F  I  L   V  G  I  I

1201 TCATCGTCAC ATCCATCAGA GGATTGCTGA TCACTCTTAC CAAGTTCTTT TATGCCATCT CTAGCAGTAA GTCCTCCAAT GTCATTGTCC TGCTATTAGC
     AGTAGCAGTG TAGGTAGTCT CCTAACGACT AGTGAGAATG GTTCAAGAAA ATACGGTAGA GATCGTCATT CAGGAGGTTA CAGTAACAGG ACGATAATCG
 352   I  V  T   S  I  R   G  L  L  I   T  L  T   K  F  F   Y  A  I  S   S  S  K   S  S  N   V  I  V  L   L  L  A

1301 ACAGATAATG GGCATGTACT TTGTCTCCTC TGTGCTGCTG ATCCGAATGA GTATGCCTTT AGAATACCGC ACCATAATCA CTGAAGTCCT TGGAGAACTG
     TGTCTATTAC CCGTACATGA AACAGAGGAG ACACGACGAC TAGGCTTACT CATACGGAAA TCTTATGGCG TGGTATTAGT GACTTCAGGA ACCTCTTGAC
 385 Q  I  M   G  M  Y  F   V  S  S   V  L  L   I  R  M  S   M  P  L   E  Y  R   T  I  I  T   E  V  L   G  E  L

1401 CAGTTCAACT TCTATCACCG TTGGTTTGAT GTGATCTTCC TGGTCAGCGC TCTCTCTAGC ATACTCTTCC TCTATTTGGC TCACAAACAG GCACCAGAGA
     GTCAAGTTGA AGATAGTGGC AACCAAACTA CACTAGAAGG ACCAGTCGCG AGAGAGATCG TATGAGAAGG AGATAAACCG AGTGTTTGTC CGTGGTCTCT
 418 Q  F  N  F   Y  H  R   W  F  D   V  I  F  L   V  S  A   L  S  S   I  L  F  L   Y  L  A   H  K  Q   A  P  E  K

1501 AGCAAATGGC ACCTCGAACT TAAGCCTACT ACAGACTGTT AGAGGCCAGT GGTTTCAAAA TTTAGATATA AGAGGGGGGA AAAATGGAAC CAGGGCCTGA
     TCGTTTACCG TGGAACTTGA ATTCGGATGA TGTCTGACAA TCTCCGGTCA CCAAAGTTTT AAATCTATAT TCTCCCCCCT TTTTACCTTG GTCCCGGACT
 452   Q  M  A   P  O

1601 CATTTTATAA ACAAACAAAA TGCTATGGTA GCATTTTTCA CCTTCATAGC ATACTCCTTC CCCGTCAGGT GATACTATGA CCATGAGTAG CATCAGCCAG
     GTAAAATATT TGTTTGTTTT ACGATACCAT CGTAAAAAGT GGAAGTATCG TATGAGGAAG GGGCAGTCCA CTATGATACT GGTACTCATC GTAGTCGGTC
                                                                                                    ^30867.tm.f

1701 AACATGAGAG GGAGAACTAA CTCAAGACAA TACTCAGCAG AGAGCATCCC GTGTGGATAT GAGGCTGGTG TAGAGGCGGA GAGGAGCCAA GAAACTAAAG
     TTGTACTCTC CCTCTTGATT GAGTTCTGTT ATGAGTCGTC TCTCGTAGGG CACACCTATA CTCCGACCAC ATCTCCGCCT CTCCTCGGTT CTTTGATTTC
                          ^30867.tm.p                                              ^30867.tm.r

1801 GTGAAAAATA CACTGGAACT CTGGGGCAAG ACATGTCTAT GGTAGCTGAG CCAAACACGT AGGATTTCCG TTTTAAGGTT CACATGGAAA AGGTTATAGC
     CACTTTTTAT GTGACCTTGA GACCCCGTTC TGTACAGATA CCATCGACTC GGTTTGTGCA TCCTAAAGGC AAAATTCCAA GTGTACCTTT TCCAATATCG

1901 TTTGCCTTGA GATTGACTCA TTAAAATCAG AGACTGTAAC AAAAAAAAAA AAAAAAAAAA AGGGCGGCCG CGACTCTAGA GTCGACCTGC AGAAGCTTGG
     AAACGGAACT CTAACTGAGT AATTTTAGTC TCTGACATTG TTTTTTTTTT TTTTTTTTTT TCCCGCCGGC GCTGAGATCT CAGCTGGACG TCTTCGAACC

2001 CCGCCATGGC CCAACTTGTT TATTGCAGCT TATAATG
     GGCGGTACCG GGTTGAACAA ATAACGTCGA ATATTAC
```

FIGURE 1B

```
          <subunit 1 of 1, 455 aa, 1 stop
          <MW: 52917, pI: 9.50, NX(S/T): 3
          <     1       10        20        30        40        50        60        70
                |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |
SEQ. ID NO:2      MSFLIDSSIMITSQILFFGFGWLFFMRQLFKDYEIRQYVVQVIFSVTFAFSCTMFELIIFEILGVLNSSS

          <    71      80        90        100       110       120       130       140
          <    |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |
               RYFHWKMNLCVILLILVFMVPFYIGYFIVSNIRLLHKQRLLFSCLLWLTFMYFFWKLGDPFPILSPKHGI

          <    141     150       160       170       180       190       200       210
          <    |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |
               LSIEQLISRVGVIGVTLMALLSGFGAVNCPYTYMSYFLRNVTDTDILALERRLLQTMDMIISKKKRMAMA

          <    211     220       230       240       250       260       270       280
          <    |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |
               RRTMFQKGEVHNKPSGFWGMIKSVTTSASGSENLTLIQQEVDALEELSRQLFLETADLYATKERIEYSKT

          <    281     290       300       310       320       330       340       350
          <    |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |
               FKGKYFNFLGYFFSIYCVWKIFMATINIVFDRVGKTDPVTRGIEITVNYLGIQFDVKFWSQHISFILVGI

          <    351     360       370       380       390       400       410       420
          <    |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |
               IIVTSIRGLLITLTKFFYAISSSKSSNVIVLLLAQIMGMYFVSSVLLIRMSMPLEYRTIITEVLGELQFN

          <    421     430       440       450
          <    |   |   |   |   |   |   |   |
               FYHRWFDVIFLVSALSSILFLYLAHKQAPEKQMAP
```

FIGURE 2A

Signal peptide:
1-23 MSFLIDSSIMITSQILFFGFGWL

Potential transmembrane domain:

37-55	QYVVQVIFSVTFAFSCTMF
81-102	VILLILVFMVPFYIGYFIVSNI
150-168	VGVIGVTLMALLSGFGAVN
288-311	FLGYFFSIYCVWKIFMATINIVFD
338-356	FWSQHISFILVGIIIVTSI
375-398	SSNVIVLLLAQIMGMYFVSSVLLI
425-444	WFDVIFLVSALSSILFLYLA

N-glycosylation site.

67	NSSS
180	NVTD
243	NLTL

Eukaryotic cobalamin-binding proteins
151 GVIGVTLMAL

FIGURE 2B

DNA14472:

```
>< /usr/seqdb2/sst/DNA/Dnaseqs.full/ss.DNA14472 (553 bases)
< good sequence: 1-553 (553 bases)
< insert: 1-407 (407 bases), 5 regions found
<    3' pSST-AMY.1 + lk: 408-1634, 145 matches (99%), 93 consec
<    cDNA linker: 408-1634, 20 matches (100%), 20 consec
<    cDNA linker: 408-6633, 16 matches (80%), 12 consec
<    cDNA linker: 422-6633, 12 matches (80%), 7 consec, 3 gaps
<    Amylase: 462-1688, 91 matches (98%), 52 consec
CATGGGAAGTGGAGCCGGAGCCTTCCTTACACTCGCCATGAGTTTCCTCATCGACTCCAG
CATCATGATTACCTCCCNGANACTATTTTTTGGATTTGGGTGGCTTTTCTTCNGCGCCAA
TGTTTAAAGACTATGAGATACGTCAGTATGTTGTACNGGTGATCTTCTCCGTGACGTTTG
CCATTTCTTGCACCATGTTTGAGCTCATCATCTTTGAAATCTTNGGAGTATTGAATAGCA
GCTCCCGTTATTTTCACTGGAAAATGAACCTGTGTGTAATTCTGCTGATCCTGGTTNTCA
TGGTGCCTTTTTACATTGGCTATTTTATTGTGAGCAATATCCGACTACTGCATAAACAAC
GACTGCTTTTTTCCTGTCTCTTATGGCTGACCTTTATGTATTTCCAG
```

SEQ. ID NO: 3

DNA15846:

```
DNA15846 [ABI300.14, LIB97, YST108, Human salivary Gland] [rev comp]
good sequence: 289-920 (632 bases)
insert: 1-745 (457 bases), 5 regions found
    3' pSST-AMY.1 + lk: 746-1634, 162 matches (97%), 93 consec, 2 gaps
    cDNA linker: 746-1634, 20 matches (100%), 20 consec
    cDNA linker: 746-6633, 16 matches (80%), 12 consec
    cDNA linker: 760-6633, 12 matches (80%), 7 consec, 3 gaps
    Amylase: 800-1688, 108 matches (96%), 39 consec, 2 gaps
GTGTTGCCCTTGGGGAGGGGAAGGGGAGCCNGGCCCTTTCCTAAAATTTGGCCAAGGGTT
TCTTTNTTGAATTCCGGGTTNNGNATACCTTCCCAGAAAATATTTTTTGGATTTGGGGTA
GNTTTTTTTCATGCGCCAATTGTTTAAAGACTATGAGATACGTCAGTATGTTGTACAGGT
GATNTTNTCCGTGACGTTTGCATTTTCTTGCACCATGTTTGAGCTCATCATNTTTGAAAT
NTTAGGAGTATTGAATAGCAGCTCCCGTTATTTTCACTGGAAAATGAACCTGTGTGTAAT
TCTGCTGATCCTGGTTTTCATGGTGCCTTTTTACATTGGCTATTTTATTGTGAGCAATAT
CCGACTACTGCATAAACAACGACTGCTTTTTTCCTGTCTNTTATGGCTGACCTTTATGTA
TTTNTTNTGGAAANTAGGAGATCCCTTTCCCATTCTC
```

SEQ. ID NO: 4

FIGURE 3

> DNA17411 [Min]
> 0 Sites [No Sites]
> length: 637 bp (circular)

```
     1 CATGGGAAGT GGAGCCGGAG CCTTCCTTAC ACTCGCCATG AGTTTCCTCA TCGACTCCAG CATCATGATT ACCTCCCAGA TACTATTTTT TGGATTTGGG
       GTACCCTTCA CCTCGGCCTC GGAAGGAATG TGAGCGGTAC TCAAAGGAGT AGCTGAGGTC GTAGTACTAA TGGAGGGTCT ATGATAAAAA ACCTAAACCC

   101 TGGCTTTTCT TCATGCGCCA ATTGTTTAAA GACTATGAGA TACGTCAGTA TGTTGTACAG GTGATCTTCT CCGTGACGTT TGCATTTTCT TGCACCATGT
       ACCGAAAAGA AGTACGCGGT TAACAAATTT CTGATACTCT ATGCAGTCAT ACAACATGTC CACTAGAAGA GGCACTGCAA ACGTAAAAGA ACGTGGTACA

   201 TTGAGCTCAT CATCTTTGAA ATCTTAGGAG TATTGAATAG CAGCTCCCGT TATTTTCACT GGAAAATGAA CCTGTGTGTA ATTCTGCTGA TCCTGGTTTT
       AACTCGAGTA GTAGAAACTT TAGAATCCTC ATAACTTATC GTCGAGGGCA ATAAAAGTGA CCTTTTACTT GGACACACAT TAAGACGACT AGGACCAAAA

   301 CATGGTGCCT TTTTACATTG GCTATTTTAT TGTGAGCAAT ATCCGACTAC TGCATAAACA ACGACTGCTT TTTTCCTGTC TCTTATGGCT GACCTTTATG
       GTACCACGGA AAAATGTAAC CGATAAAATA ACACTCGTTA TAGGCTGATG ACGTATTTGT TGCTGACGAA AAAAGGACAG AGAATACCGA CTGGAAATAC

   401 TATTTCTTCT GGAAACTAGG AGATCCCTTT CCCATTCTCA GCCCAAAACA TGGGATCTTA TCCNTANAGN ACNAGCTNNT NANCCGGGTT GGTGTGATTG
       ATAAAGAAGA CCTTTGATCC TCTAGGGAAA GGGTAAGAGT CGGGTTTTGT ACCCTAGAAT AGGNATNTCN TGNTCGANNA NTNGGCCCAA CCACACTAAC

   501 GAGTGACTCT CATGGCTCTT CTTTCTGGAT TTGGTGCTGT CAACTGCCCA TACACTTACA TGTCTTACTT CCTCAGGAAT GTGACTGACA CGGATATTCT
       CTCACTGAGA GTACCGAGAA GAAAGACCTA AACCACGACA GTTGACGGGT ATGTGAATGT ACAGAATGAA GGAGTCCTTA CACTGACTGT GCCTATAAGA

   601 AGCCCTGGAA CGGCGACTGC TGCAAACCAT GGATATG
       TCGGGACCTT GCCGCTGACG ACGTTTGGTA CCTATAC
```

SEQ. ID NO:8

FIGURE 4

587557

PROVISIONAL APPLICATION COVER SHEET

This is a request for filing a PROVISIONAL APPLICATION under 37 CFR 1.53(b)(2)

"EXPRESS MAIL" MAILING LABEL
Number: EL029353477US Date of Deposit: May 15, 1998
I hereby certify that this paper or fee is being deposited with the United States Postal Service "Express Mail Post Office to Addressee" service under 37 C.F.R. 1.10 on the date indicated above and is addressed to the Box Provisional Patent Application, Assistant Commissioner for Patents, Washington, D.C. 20231.
Typed or Printed Name: Geody Domingo
Signed:

05/15/98

Docket Number:	P-66136/WHD/DAV (G-tech Docket: PR1381)	Type a plus sign (+) inside this box →	+

INVENTOR(s)/APPLICANT(s)

LAST NAME	FIRST NAME	MIDDLE INITIAL	RESIDENCE (CITY AND EITHER STATE OR FOREIGN COUNTRY)
WOOD	WILLIAM	I.	San Mateo, California

TITLE OF THE INVENTION (280 characters max)

NOVEL POLYPEPTIDES HAVING SEQUENCE IDENTITY WITH IL-17 AND NUCLEIC ACIDS ENCODING THE SAME

CORRESPONDENCE ADDRESS

Walter H. Dreger, Registration No. 24, 190
FLEHR HOHBACH TEST ALBRITTON & HERBERT LLP, 4 Embarcadero Center, Suite 3400, San Francisco

STATE	CA	ZIP CODE	94111	COUNTRY	US

ENCLOSED APPLICATION PARTS (check all that apply)

X	Specification	Number of Pages	39		Small Entity Statement
X	Drawings	Number of Sheets	5		Other (specify)

METHOD OF PAYMENT OF FILING FEES FOR THIS PROVISIONAL APPLICATION FOR PATENT (check one)

X	A check or money order is enclosed to cover the Provisional filing fees		PROVISIONAL FILING FEE AMOUNT ($)	$150
X	The Commissioner is hereby authorized to charge any underpayment or overpayment to Deposit Account Number.	06-1300 (Order No. P-66136/WHD/DAV)		

The invention was made by an agency of the United States Government or under a contract with an agency of the United States Government.

X No

☐ Yes, the name of the U.S. Government Agency and the Government contract number are: _____

Respectfully submitted, rgn 39,054

SIGNATURE: Walter H. Dreger Date 15 May 1998

TYPED or PRINTED NAME Walter H. Dreger REGISTRATION NO (if appropriate) 24,190

☐ Additional inventors are being named on separately numbered sheets attached hereto.

USE ONLY FOR FILING A PROVISIONAL APPLICATION FOR PATENT

Burden Hour Statement: This form is estimated to take .2 hours to complete. Times will vary depending upon the needs of the individual case. Any comments on the amount of time you are required to complete this form should be sent to the Chief Information Officer, Assistant Commissioner for Patents, Washington, DC 20231 DO NOT SEND FEES OR COMPLETED FORMS TO THIS ADDRESS. SEND TO: Box Provisional Application, Assistant Commissioner for Patents, Washington, D.C. 20231.

NOVEL POLYPEPTIDES HAVING SEQUENCE IDENTITY WITH IL-17 AND NUCLEIC ACIDS ENCODING THE SAME

FIELD OF THE INVENTION

The present invention relates generally to the identification and isolation of novel DNA and to the recombinant production of novel polypeptides having sequence identity with the cytokine IL-17, designated herein as "PRO1031" polypeptides.

BACKGROUND OF THE INVENTION

It has been reported that the cytokine interleukin 17 (IL-17) stimulates epithelial, endothelial, and fibroblastic cells to secrete cytokines such as IL-6, IL-8, and granulocyte-colony-stimulating factor, as well as prostaglandin E2. Moreover, it has been shown that when cultured in the presence of IL-17, fibroblasts could sustain proliferation of CD34+ preferential maturation into neutrophils. Thus it has been suggested that IL-17 constitutes an early initiator of the T cell-dependent inflammatory reaction and/or an element of the cytokine network that bridges the immune system to hematopoiesis. See, Yao, et al., J. Immunol., 155(12):5483-5486 (1995); Fossiez, et al., J. Exp. Med., 183(6):2593-2603 (1996); Kennedy, et al., J. Interferon Cytokine Res., 16(8):611-617 (1996). Thus, proteins related to IL-17 are of interest.

More generally, all novel proteins are of interest. Extracellular proteins play an important role in the formation, differentiation and maintenance of multicellular organisms. The fate of many individual cells, e.g., proliferation, migration, differentiation, or interaction with other cells, is typically governed by information received from other cells and/or the immediate environment. This information is often transmitted by secreted polypeptides (for instance, mitogenic factors, survival factors, cytotoxic factors, differentiation factors, neuropeptides, and hormones) which are, in turn, received and interpreted by diverse cell receptors or membrane-bound proteins. These secreted polypeptides or signaling molecules normally pass through the cellular secretory pathway to reach their site of action in the extracellular environment.

Secreted proteins have various industrial applications, including pharmaceuticals, diagnostics, biosensors and bioreactors. Most protein drugs available at present, such as thrombolytic agents, interferons, interleukins, erythropoietins, colony stimulating factors, and various other cytokines, are secretory proteins. Their receptors, which are membrane proteins, also have potential as therapeutic or diagnostic agents.

Efforts are being undertaken by both industry and academia to identify new, native secreted proteins, particularly those related to IL-17. Many efforts are focused on the screening of mammalian recombinant DNA libraries to identify the coding sequences for novel secreted proteins. Examples of screening methods and techniques are described in the literature [see, for example, Klein et al., Proc. Natl. Acad. Sci., 93:7108-7113 (1996); U.S. Patent No. 5,536,637)]. The results of such efforts are presented herein.

SUMMARY OF THE INVENTION

Applicants have identified a cDNA clone that encodes a novel polypeptide having sequence identity with IL-17, wherein the polypeptide is designated in the present application as "PRO1031".

In one embodiment, the invention provides an isolated nucleic acid molecule comprising DNA encoding a PRO1031 polypeptide. In one aspect, the isolated nucleic acid comprises DNA encoding the PRO1031 polypeptide having amino acid residues 1 through 180 of Figure 2 (SEQ ID NO:2), or is complementary to such encoding nucleic acid sequence, and remains stably bound to it under at least moderate, and optionally, under high stringency conditions. The isolated nucleic acid sequence may comprise the cDNA insert of the vector deposited on 14 May 1998 with the ATCC as DNA59294-1381 which includes the nucleotide sequence encoding PRO1031.

In another embodiment, the invention provides a vector comprising DNA encoding a PRO1031 polypeptide. A host cell comprising such a vector is also provided. By way of example, the host cells may be CHO cells, *E. coli*, or yeast. A process for producing PRO1031 polypeptides is further provided and comprises culturing host cells under conditions suitable for expression of PRO1031 and recovering PRO1031 from the cell culture.

In another embodiment, the invention provides isolated PRO1031 polypeptide. In particular, the invention provides isolated native sequence PRO1031 polypeptide, which in one embodiment, includes an amino acid sequence comprising residues 1 through 180 of Figure 2 (SEQ ID NO:2). Optionally, the PRO1031 polypeptide is obtained or is obtainable

by expressing the polypeptide encoded by the cDNA insert of the vector deposited on 14 May 1998 with the ATCC as DNA59294-1381.

In another embodiment, the invention provides chimeric molecules comprising a PRO1031 polypeptide fused to a heterologous polypeptide or amino acid sequence. An example of such a chimeric molecule comprises a PRO1031 polypeptide fused to an epitope tag sequence or a Fc region of an immunoglobulin.

In another embodiment, the invention provides an antibody which specifically binds to a PRO1031 polypeptide. Optionally, the antibody is a monoclonal antibody.

BRIEF DESCRIPTION OF THE DRAWINGS

Figure 1 shows a nucleotide sequence (SEQ ID NO:1) containing the nucleotide sequence of a native sequence PRO1031 cDNA (nucleotides 42-581 of SEQ ID NO:1), wherein the nucleotide sequence (SEQ ID NO:1) is a clone designated herein as "UNQ516" and/or "DNA59294-1381". Also presented is the position of the initiator methionine residue, circled, and the stop codon, circled. The complementary strand and deduced amino acid sequence are also shown.

Figure 2 shows the amino acid sequence (SEQ ID NO:2) derived from nucleotides 42-581 of SEQ ID NO:1. Also shown in Figure 2 are the signal peptide, an N-glycosylation site, a region having sequence identity with IL-17, the molecular weight, and approximate pI.

Figures 3A-3C show an alignment of nucleotide sequences from a variety of expressed sequence tags as well as a consensus nucleotide sequence derived therefrom designated "DNA-from dna" shown as "<consen01>".

DETAILED DESCRIPTION OF THE PREFERRED EMBODIMENTS

I. Definitions

The terms "PRO1031 polypeptide" and "PRO1031" when used herein encompass native sequence PRO1031 and PRO1031 polypeptide variants (which are further defined herein). The PRO1031 polypeptides may be isolated from a variety of sources, such as from human tissue types or from another source, or prepared by recombinant or synthetic methods.

A "native sequence PRO1031 polypeptide" comprises a polypeptide having the same amino acid sequence as a PRO1031 polypeptide derived from nature. Such native sequence PRO1031 polypeptide can be isolated from nature or can be produced by recombinant or

synthetic means. The term "native sequence PRO1031 polypeptide" specifically encompasses naturally-occurring truncated or secreted forms of a PRO1031 polypeptide (*e.g.*, soluble forms containing for instance, an extracellular domain sequence), naturally-occurring variant forms (*e.g.*, alternatively spliced forms) and naturally-occurring allelic variants of a PRO1031 polypeptide. In one embodiment of the invention, the native sequence PRO1031 polypeptide is a full-length or mature native sequence PRO1031 polypeptide comprising amino acids 1 or 21 through 180 of Figure 2 (SEQ ID NO:2). Optionally, the PRO1031 polypeptide is obtained or obtainable by expressing the polypeptide encoded by the cDNA insert of the vector DNA59294-1381 deposited on 14 May 1998 with the ATCC.

It will be understood that any transmembrane domain identified for the PRO1031 polypeptide of the present invention is identified pursuant to criteria routinely employed in the art for identifying that type of hydrophobic domain. The exact boundaries of a transmembrane domain may vary but most likely by no more than about 5 amino acids at either end of the domain as initially identified.

"PRO1031 variant" means an active PRO1031 polypeptide as defined below having at least about 80% amino acid sequence identity with the PRO1031 polypeptide having the deduced amino acid sequence shown in Figure 2 (SEQ ID NO:2) for a full-length or mature native sequence PRO1031 polypeptide. Such PRO1031 polypeptide variants include, for instance, PRO1031 polypeptides wherein one or more amino acid residues are added, substituted or deleted, at the N- or C-terminus or within the sequence of Figure 2 (SEQ ID NO:2). Ordinarily, a PRO1031 polypeptide variant will have at least about 80% amino acid sequence identity, preferably at least about 85% amino acid sequence identity, more preferably at least about 90% amino acid sequence identity and even more preferably at least about 95% amino acid sequence identity with the amino acid sequence of Figure 2 (SEQ ID NO:2), with or without the signal peptide. The variants provided herein exclude the polypeptides and nucleic acids described herein which the novel polypeptides have identity with and which are already known in the art.

"Percent (%) amino acid sequence identity" with respect to the PRO1031 amino acid sequences identified herein is defined as the percentage of amino acid residues in a candidate sequence that are identical with the amino acid residues in a PRO1031 polypeptide sequence, after aligning the sequences and introducing gaps, if necessary, to achieve the maximum percent sequence identity, and not considering any conservative substitutions as part of the sequence identity. Alignment for purposes of determining percent amino acid sequence

identity can be achieved in various ways that are within the skill in the art, for instance, using publicly available computer software such as ALIGN or Megalign (DNASTAR) software. Those skilled in the art can determine appropriate parameters for measuring alignment, including any algorithms needed to achieve maximal alignment over the full length of the sequences being compared. Similarity can be determined using the same programs.

"Percent (%) nucleic acid sequence identity" with respect to the PRO1031 sequence identified herein is defined as the percentage of nucleotides in a candidate sequence that are identical with the nucleotides in the PRO1031 sequence, after aligning the sequences and introducing gaps, if necessary, to achieve the maximum percent sequence identity. Alignment for purposes of determining percent nucleic acid sequence identity can be achieved in various ways that are within the skill in the art, for instance, using publicly available computer software such as ALIGN or Megalign (DNASTAR) software. Those skilled in the art can determine appropriate parameters for measuring alignment, including any algorithms needed to achieve maximal alignment over the full length of the sequences being compared. Similarity can be determined using the same programs.

Most of the WU-BLAST-2 search parameters as used herein were set to the default values. The adjustable parameters were set with the following values: overlap span = 1, overlap fraction = 0.125, word threshold (T) = 11, and scoring matrix = BLOSUM62. The HSP S and HSP S2 parameters, which are dynamic values used by BLAST-2, are established by the program itself depending upon the composition of the sequence of interest and composition of the database against which the sequence is being searched. However, the values may be adjusted to increase sensitivity. A % sequence identity value is determined by the fraction of matching identical residues divided by the total number of residues of the "longer" sequence. The "longer" sequence is the one having the most actual residues in the aligned region (gaps introduced by WU-BLAST-2 to maximize the alignment score are ignored).

The term "epitope tagged" where used herein refers to a chimeric polypeptide comprising a PRO1031 polypeptide, or domain sequence thereof, fused to a "tag polypeptide". The tag polypeptide has enough residues to provide an epitope against which an antibody may be made, or which can be identified by some other agent, yet is short enough such that it does not interfere with the activity of the PRO1031 polypeptide. The tag polypeptide preferably is also fairly unique so that the antibody does not substantially cross-react with other epitopes. Suitable tag polypeptides generally have at least six amino acid

residues and usually between about 8 to about 50 amino acid residues (preferably, between about 10 to about 20 residues).

"Isolated," when used to describe the various polypeptides disclosed herein, means polypeptide that has been identified and separated and/or recovered from a component of its natural environment. Contaminant components of its natural environment are materials that would typically interfere with diagnostic or therapeutic uses for the polypeptide, and may include enzymes, hormones, and other proteinaceous or non-proteinaceous solutes. In preferred embodiments, the polypeptide will be purified (1) to a degree sufficient to obtain at least 15 residues of N-terminal or internal amino acid sequence by use of a spinning cup sequenator, or (2) to homogeneity by SDS-PAGE under non-reducing or reducing conditions using Coomassie blue or, preferably, silver stain. Isolated polypeptide includes polypeptide *in situ* within recombinant cells, since at least one component of the PRO1031 polypeptide natural environment will not be present. Ordinarily, however, isolated polypeptide will be prepared by at least one purification step.

An "isolated" PRO1031 polypeptide-encoding nucleic acid molecule is a nucleic acid molecule that is identified and separated from at least one contaminant nucleic acid molecule with which it is ordinarily associated in the natural source of the PRO1031 polypeptide-encoding nucleic acid. An isolated PRO1031 polypeptide-encoding nucleic acid molecule is other than in the form or setting in which it is found in nature. Isolated PRO1031 polypeptide-encoding nucleic acid molecules therefore are distinguished from the PRO1031 polypeptide-encoding nucleic acid molecule as it exists in natural cells. However, an isolated PRO1031 polypeptide-encoding nucleic acid molecule includes PRO1031 polypeptide-encoding nucleic acid molecules contained in cells that ordinarily express PRO1031 polypeptide where, for example, the nucleic acid molecule is in a chromosomal location different from that of natural cells.

The term "control sequences" refers to DNA sequences necessary for the expression of an operably linked coding sequence in a particular host organism. The control sequences that are suitable for prokaryotes, for example, include a promoter, optionally an operator sequence, and a ribosome binding site. Eukaryotic cells are known to utilize promoters, polyadenylation signals, and enhancers.

Nucleic acid is "operably linked" when it is placed into a functional relationship with another nucleic acid sequence. For example, DNA for a presequence or secretory leader is operably linked to DNA for a polypeptide if it is expressed as a preprotein that participates

in the secretion of the polypeptide; a promoter or enhancer is operably linked to a coding sequence if it affects the transcription of the sequence; or a ribosome binding site is operably linked to a coding sequence if it is positioned so as to facilitate translation. Generally, "operably linked" means that the DNA sequences being linked are contiguous, and, in the case of a secretory leader, contiguous and in reading phase. However, enhancers do not have to be contiguous. Linking is accomplished by ligation at convenient restriction sites. If such sites do not exist, the synthetic oligonucleotide adaptors or linkers are used in accordance with conventional practice.

The term "antibody" is used in the broadest sense and specifically covers single anti-PRO1031 polypeptide monoclonal antibodies (including agonist, antagonist, and neutralizing antibodies) and anti-PRO1031 antibody compositions with polyepitopic specificity. The term "monoclonal antibody" as used herein refers to an antibody obtained from a population of substantially homogeneous antibodies, *i.e.*, the individual antibodies comprising the population are identical except for possible naturally-occurring mutations that may be present in minor amounts.

"Active" or "activity" for the purposes herein refers to form(s) of PRO1031 which retain the biologic and/or immunologic activities of native or naturally-occurring PRO1031 polypeptide.

The terms "treating", "treatment" and "therapy" as used herein refer to curative therapy, prophylactic therapy, and preventative therapy.

The term "mammal" as used herein refers to any mammal classified as a mammal, including humans, cows, horses, dogs and cats. In a preferred embodiment of the invention, the mammal is a human.

II. Compositions and Methods of the Invention

A. Full-length PRO1031 Polypeptide

The present invention provides newly identified and isolated nucleotide sequences encoding polypeptides referred to in the present application as PRO1031. In particular, Applicants have identified and isolated cDNA encoding a PRO1031 polypeptide, as disclosed in further detail in the Examples below. Using BLAST and FastA sequence alignment computer programs, Applicants found that various portions of the PRO1031 polypeptide have sequence identity with IL-17 and CTLA-8. Accordingly, it is presently believed that

PRO1031 polypeptide disclosed in the present application is a newly identified member of the cytokine family and thus may be involved in inflammation and/or the immune system.

B. PRO1031 Variants

In addition to the full-length native sequence PRO1031 polypeptide described herein, it is contemplated that PRO1031 variants can be prepared. PRO1031 variants can be prepared by introducing appropriate nucleotide changes into the PRO1031-encoding DNA, or by synthesis of the desired PRO1031 polypeptide. Those skilled in the art will appreciate that amino acid changes may alter post-translational processes of the PRO1031 polypeptide, such as changing the number or position of glycosylation sites or altering the membrane anchoring characteristics.

Variations in the native full-length sequence PRO1031 or in various domains of the PRO1031 polypeptide described herein, can be made, for example, using any of the techniques and guidelines for conservative and non-conservative mutations set forth, for instance, in U.S. Patent No. 5,364,934. Variations may be a substitution, deletion or insertion of one or more codons encoding the PRO1031 polypeptide that results in a change in the amino acid sequence of the PRO1031 polypeptide as compared with the native sequence PRO1031. Optionally the variation is by substitution of at least one amino acid with any other amino acid in one or more of the domains of the PRO1031 polypeptide. Guidance in determining which amino acid residue may be inserted, substituted or deleted without adversely affecting the desired activity may be found by comparing the sequence of the PRO1031 polypeptide with that of homologous known protein molecules and minimizing the number of amino acid sequence changes made in regions of high homology. Amino acid substitutions can be the result of replacing one amino acid with another amino acid having similar structural and/or chemical properties, such as the replacement of a leucine with a serine, i.e., conservative amino acid replacements. Insertions or deletions may optionally be in the range of 1 to 5 amino acids. The variation allowed may be determined by systematically making insertions, deletions or substitutions of amino acids in the sequence and testing the resulting variants for activity in any of the *in vitro* assays described in the Examples below.

The variations can be made using methods known in the art such as oligonucleotide-mediated (site-directed) mutagenesis, alanine scanning, and PCR mutagenesis. Site-directed

mutagenesis [Carter et al., Nucl. Acids Res., 13:4331 (1986); Zoller et al., Nucl. Acids Res., 10:6487 (1987)], cassette mutagenesis [Wells et al., Gene, 34:315 (1985)], restriction selection mutagenesis [Wells et al., Philos. Trans. R. Soc. London SerA, 317:415 (1986)] or other known techniques can be performed on the cloned DNA to produce the PRO1031-encoding variant DNA.

Scanning amino acid analysis can also be employed to identify one or more amino acids along a contiguous sequence. Among the preferred scanning amino acids are relatively small, neutral amino acids. Such amino acids include alanine, glycine, serine, and cysteine. Alanine is typically a preferred scanning amino acid among this group because it eliminates the side-chain beyond the beta-carbon and is less likely to alter the main-chain conformation of the variant. Alanine is also typically preferred because it is the most common amino acid. Further, it is frequently found in both buried and exposed positions [Creighton, The Proteins, (W.H. Freeman & Co., N.Y.); Chothia, J. Mol. Biol., 150:1 (1976)]. If alanine substitution does not yield adequate amounts of variant, an isoteric amino acid can be used.

C. Modifications of PRO1031

Covalent modifications of PRO1031 polypeptides are included within the scope of this invention. One type of covalent modification includes reacting targeted amino acid residues of a PRO1031 polypeptide with an organic derivatizing agent that is capable of reacting with selected side chains or the N- or C- terminal residues of a PRO1031 polypeptide. Derivatization with bifunctional agents is useful, for instance, for crosslinking PRO1031 to a water-insoluble support matrix or surface for use in the method for purifying anti-PRO1031 antibodies, and vice-versa. Commonly used crosslinking agents include, *e.g.*, 1,1-bis(diazo-acetyl)-2-phenylethane, glutaraldehyde, N-hydroxysuccinimide esters, for example, esters with 4-azidosalicylic acid, homobifunctional imidoesters, including disuccinimidyl esters such as 3,3'-dithiobis(succinimidylpropionate), bifunctional maleimides such as bis-N-maleimido-1,8-octane and agents such as methyl-3-[(p-azidophenyl)dithio]propioimidate.

Other modifications include deamidation of glutaminyl and asparaginyl residues to the corresponding glutamyl and aspartyl residues, respectively, hydroxylation of proline and lysine, phosphorylation of hydroxyl groups of seryl or threonyl residues, methylation of the α-amino groups of lysine, arginine, and histidine side chains [T.E. Creighton, Proteins: Structure and Molecular Properties, W.H. Freeman & Co., San Francisco, pp. 79-86

(1983)], acetylation of the N-terminal amine, and amidation of any C-terminal carboxyl group.

Another type of covalent modification of the PRO1031 polypeptide included within the scope of this invention comprises altering the native glycosylation pattern of the polypeptide. "Altering the native glycosylation pattern" is intended for purposes herein to mean deleting one or more carbohydrate moieties found in native sequence PRO1031 polypeptide, and/or adding one or more glycosylation sites that are not present in the native sequence PRO1031 polypeptide.

Addition of glycosylation sites to PRO1031 polypeptides may be accomplished by altering the amino acid sequence thereof. The alteration may be made, for example, by the addition of, or substitution by, one or more serine or threonine residues to the native sequence PRO1031 polypeptide (for O-linked glycosylation sites). The PRO1031 amino acid sequence may optionally be altered through changes at the DNA level, particularly by mutating the DNA encoding the PRO1031 polypeptide at preselected bases such that codons are generated that will translate into the desired amino acids.

Another means of increasing the number of carbohydrate moieties on the PRO1031 polypeptide is by chemical or enzymatic coupling of glycosides to the polypeptide. Such methods are described in the art, e.g., in WO 87/05330 published 11 September 1987, and in Aplin and Wriston, CRC Crit. Rev. Biochem., pp. 259-306 (1981).

Removal of carbohydrate moieties present on the PRO1031 polypeptide may be accomplished chemically or enzymatically or by mutational substitution of codons encoding for amino acid residues that serve as targets for glycosylation. Chemical deglycosylation techniques are known in the art and described, for instance, by Hakimuddin, et al., Arch. Biochem. Biophys., 259:52 (1987) and by Edge et al., Anal. Biochem., 118:131 (1981). Enzymatic cleavage of carbohydrate moieties on polypeptides can be achieved by the use of a variety of endo- and exo-glycosidases as described by Thotakura et al., Meth. Enzymol., 138:350 (1987).

Another type of covalent modification of PRO1031 comprises linking the PRO1031 polypeptide to one of a variety of nonproteinaceous polymers, *e.g.*, polyethylene glycol, polypropylene glycol, or polyoxyalkylenes, in the manner set forth in U.S. Patent Nos. 4,640,835; 4,496,689; 4,301,144; 4,670,417; 4,791,192 or 4,179,337.

PRO1031 polypeptides of the present invention may also be modified in a way to form chimeric molecules comprising a PRO1031 polypeptide fused to another, heterologous

polypeptide or amino acid sequence. In one embodiment, such a chimeric molecule comprises a fusion of a PRO1031 polypeptide with a tag polypeptide which provides an epitope to which an anti-tag antibody can selectively bind. The epitope tag is generally placed at the amino- or carboxyl- terminus of the PRO1031 polypeptide. The presence of such epitope-tagged forms of a PRO1031 polypeptide can be detected using an antibody against the tag polypeptide. Also, provision of the epitope tag enables the PRO1031 polypeptide to be readily purified by affinity purification using an anti-tag antibody or another type of affinity matrix that binds to the epitope tag. In an alternative embodiment, the chimeric molecule may comprise a fusion of a PRO1031 polypeptide with an immunoglobulin or a particular region of an immunoglobulin. For a bivalent form of the chimeric molecule, such a fusion could be to the Fc region of an IgG molecule.

Various tag polypeptides and their respective antibodies are well known in the art. Examples include poly-histidine (poly-his) or poly-histidine-glycine (poly-his-gly) tags; the flu HA tag polypeptide and its antibody 12CA5 [Field et al., Mol. Cell. Biol., 8:2159-2165 (1988)]; the c-myc tag and the 8F9, 3C7, 6E10, G4, B7 and 9E10 antibodies thereto [Evan et al., Molecular and Cellular Biology, 5:3610-3616 (1985)]; and the Herpes Simplex virus glycoprotein D (gD) tag and its antibody [Paborsky et al., Protein Engineering, 3(6):547-553 (1990)]. Other tag polypeptides include the Flag-peptide [Hopp et al., BioTechnology, 6:1204-1210 (1988)]; the KT3 epitope peptide [Martin et al., Science, 255:192-194 (1992)]; an α-tubulin epitope peptide [Skinner et al., J. Biol. Chem., 266:15163-15166 (1991)]; and the T7 gene 10 protein peptide tag [Lutz-Freyermuth et al., Proc. Natl. Acad. Sci. USA, 87:6393-6397 (1990)].

The PRO1031 polypeptide of the present invention may also be modified in a way to form a chimeric molecule comprising a PRO1031 polypeptide fused to a leucine zipper. Various leucine zipper polypeptides have been described in the art. See, e.g., Landschulz et al., Science 240:1759 (1988); WO 94/10308; Hoppe et al., FEBS Letters 344:1991 (1994); Maniatis et al., Nature 341:24 (1989). It is believed that use of a leucine zipper fused to a PRO1031 polypeptide may be desirable to assist in dimerizing or trimerizing soluble PRO1031 polypeptide in solution. Those skilled in the art will appreciate that the leucine zipper may be fused at either the N- or C-terminal end of the PRO1031 molecule.

D. Preparation of PRO1031

The description below relates primarily to production of PRO1031 by culturing cells transformed or transfected with a vector containing PRO1031 polypeptide encoding nucleic acid. It is, of course, contemplated that alternative methods, which are well known in the art, may be employed to prepare PRO1031 polypeptides. For instance, the PRO1031 sequence, or portions thereof, may be produced by direct peptide synthesis using solid-phase techniques [see, e.g., Stewart et al., Solid-Phase Peptide Synthesis, W.H. Freeman Co., San Francisco, CA (1969); Merrifield, J. Am. Chem. Soc., 85:2149-2154 (1963)]. *In vitro* protein synthesis may be performed using manual techniques or by automation. Automated synthesis may be accomplished, for instance, using an Applied Biosystems Peptide Synthesizer (Foster City, CA) using manufacturer's instructions. Various portions of PRO1031 polypeptides may be chemically synthesized separately and combined using chemical or enzymatic methods to produce a full-length PRO1031 polypeptide.

1. Isolation of DNA Encoding PRO1031

DNA encoding a PRO1031 polypeptide may be obtained from a cDNA library prepared from tissue believed to possess the PRO1031 mRNA and to express it at a detectable level. Accordingly, human PRO1031-encoding DNA can be conveniently obtained from a cDNA library prepared from human tissue, such as described in the Examples. The PRO1031-encoding gene may also be obtained from a genomic library or by oligonucleotide synthesis.

Libraries can be screened with probes (such as antibodies to a PRO1031 polypeptide or oligonucleotides of at least about 20-80 bases) designed to identify the gene of interest or the protein encoded by it. Screening the cDNA or genomic library with the selected probe may be conducted using standard procedures, such as described in Sambrook et al., Molecular Cloning: A Laboratory Manual (New York: Cold Spring Harbor Laboratory Press, 1989). An alternative means to isolate the gene encoding PRO1031 is to use PCR methodology [Sambrook et al., supra; Dieffenbach et al., PCR Primer:A Laboratory Manual (Cold Spring Harbor Laboratory Press, 1995)].

The Examples below describe techniques for screening a cDNA library. The oligonucleotide sequences selected as probes should be of sufficient length and sufficiently unambiguous that false positives are minimized. The oligonucleotide is preferably labeled such that it can be detected upon hybridization to DNA in the library being screened.

Methods of labeling are well known in the art, and include the use of radiolabels like ^{32}P-labeled ATP, biotinylation or enzyme labeling. Hybridization conditions, including moderate stringency and high stringency, are provided in Sambrook et al., supra.

Included in this invention are nucleic acids having at least about 70% to about 75% nucleic acid sequence identity with the nucleic acid sequence shown in Figure 1 (SEQ ID NO:1), with or without the untranslated portions and with or without the region encoding the signal peptide. Ordinarily, in another embodiment, the nucleic acids will have at least about 80% nucleic acid sequence identity, preferably at least about 85% nucleic acid sequence identity, more preferably at least about 90% nucleic acid sequence identity and even more preferably at least about 95% nucleic acid sequence identity with the nucleic acid sequence of SEQ ID NO:1, with or without the untranslated portions and with or without the region encoding the signal peptide.

Sequences identified in such library screening methods can be compared and aligned to other known sequences deposited and available in public databases such as GenBank or other private sequence databases. Sequence identity (at either the amino acid or nucleotide level) within defined regions of the molecule or across the full-length sequence can be determined through sequence alignment using computer software programs such as ALIGN, DNAstar, and INHERIT which employ various algorithms to measure homology.

Nucleic acid having protein coding sequence may be obtained by screening selected cDNA or genomic libraries using the deduced amino acid sequence disclosed herein for the first time, and, if necessary, using conventional primer extension procedures as described in Sambrook et al., supra, to detect precursors and processing intermediates of mRNA that may not have been reverse-transcribed into cDNA.

2. Selection and Transformation of Host Cells

Host cells are transfected or transformed with expression or cloning vectors described herein for PRO1031 polypeptide production and cultured in conventional nutrient media modified as appropriate for inducing promoters, selecting transformants, or amplifying the genes encoding the desired sequences. The culture conditions, such as media, temperature, pH and the like, can be selected by the skilled artisan without undue experimentation. In general, principles, protocols, and practical techniques for maximizing the productivity of cell cultures can be found in Mammalian Cell Biotechnology: a Practical Approach, M. Butler, ed. (IRL Press, 1991) and Sambrook et al., supra.

Methods of transfection are known to the ordinarily skilled artisan, for example, $CaPO_4$ and electroporation. Depending on the host cell used, transformation is performed using standard techniques appropriate to such cells. The calcium treatment employing calcium chloride, as described in Sambrook et al., supra, or electroporation is generally used for prokaryotes or other cells that contain substantial cell-wall barriers. Infection with *Agrobacterium tumefaciens* is used for transformation of certain plant cells, as described by Shaw et al., Gene, 23:315 (1983) and WO 89/05859 published 29 June 1989. For mammalian cells without such cell walls, the calcium phosphate precipitation method of Graham and van der Eb, Virology, 52:456-457 (1978) can be employed. General aspects of mammalian cell host system transformations have been described in U.S. Patent No. 4,399,216. Transformations into yeast are typically carried out according to the method of Van Solingen et al., J. Bact., 130:946 (1977) and Hsiao et al., Proc. Natl. Acad. Sci. (USA), 76:3829 (1979). However, other methods for introducing DNA into cells, such as by nuclear microinjection, electroporation, bacterial protoplast fusion with intact cells, or polycations, *e.g.*, polybrene, polyornithine, may also be used. For various techniques for transforming mammalian cells, see Keown et al., Methods in Enzymology, 185:527-537 (1990) and Mansour et al., Nature, 336:348-352 (1988).

Suitable host cells for cloning or expressing the DNA in the vectors herein include prokaryote, yeast, or higher eukaryote cells. Suitable prokaryotes include but are not limited to eubacteria, such as Gram-negative or Gram-positive organisms, for example, Enterobacteriaceae such as *E. coli*. Various *E. coli* strains are publicly available, such as *E. coli* K12 strain MM294 (ATCC 31,446); *E. coli* X1776 (ATCC 31,537); *E. coli* strain W3110 (ATCC 27,325) and K5 772 (ATCC 53,635).

In addition to prokaryotes, eukaryotic microbes such as filamentous fungi or yeast are suitable cloning or expression hosts for PRO1031-encoding vectors. *Saccharomyces cerevisiae* is a commonly used lower eukaryotic host microorganism.

Suitable host cells for the expression of glycosylated PRO1031 are derived from multicellular organisms. Examples of invertebrate cells include insect cells such as Drosophila S2 and Spodoptera Sf9, as well as plant cells. Examples of useful mammalian host cell lines include Chinese hamster ovary (CHO) and COS cells. More specific examples include monkey kidney CV1 line transformed by SV40 (COS-7, ATCC CRL 1651); human embryonic kidney line (293 or 293 cells subcloned for growth in suspension culture, Graham et al., J. Gen Virol., 36:59 (1977)); Chinese hamster ovary cells/-DHFR (CHO, Urlaub and

Chasin, Proc. Natl. Acad. Sci. USA, 77:4216 (1980)); mouse sertoli cells (TM4, Mather, Biol. Reprod., 23:243-251 (1980)); human lung cells (W138, ATCC CCL 75); human liver cells (Hep G2, HB 8065); and mouse mammary tumor (MMT 060562, ATCC CCL51). The selection of the appropriate host cell is deemed to be within the skill in the art.

3. Selection and Use of a Replicable Vector

The nucleic acid (*e.g.*, cDNA or genomic DNA) encoding the desired PRO1031 polypeptide may be inserted into a replicable vector for cloning (amplification of the DNA) or for expression. Various vectors are publicly available. The vector may, for example, be in the form of a plasmid, cosmid, viral particle, or phage. The appropriate nucleic acid sequence may be inserted into the vector by a variety of procedures. In general, DNA is inserted into an appropriate restriction endonuclease site(s) using techniques known in the art. Vector components generally include, but are not limited to, one or more of a signal sequence, an origin of replication, one or more marker genes, an enhancer element, a promoter, and a transcription termination sequence. Construction of suitable vectors containing one or more of these components employs standard ligation techniques which are known to the skilled artisan.

The desired PRO1031 polypeptide may be produced recombinantly not only directly, but also as a fusion polypeptide with a heterologous polypeptide, which may be a signal sequence or other polypeptide having a specific cleavage site at the N-terminus of the mature protein or polypeptide. In general, the signal sequence may be a component of the vector, or it may be a part of the PRO1031-encoding DNA that is inserted into the vector. The signal sequence may be a prokaryotic signal sequence selected, for example, from the group of the alkaline phosphatase, penicillinase, lpp, or heat-stable enterotoxin II leaders. For yeast secretion the signal sequence may be, *e.g.*, the yeast invertase leader, alpha factor leader (including *Saccharomyces* and *Kluyveromyces* α-factor leaders, the latter described in U.S. Patent No. 5,010,182), or acid phosphatase leader, the *C. albicans* glucoamylase leader (EP 362,179 published 4 April 1990), or the signal described in WO 90/13646 published 15 November 1990. In mammalian cell expression, mammalian signal sequences may be used to direct secretion of the protein, such as signal sequences from secreted polypeptides of the same or related species, as well as viral secretory leaders.

Both expression and cloning vectors contain a nucleic acid sequence that enables the vector to replicate in one or more selected host cells. Such sequences are well known for

a variety of bacteria, yeast, and viruses. The origin of replication from the plasmid pBR322 is suitable for most Gram-negative bacteria, the 2μ plasmid origin is suitable for yeast, and various viral origins (SV40, polyoma, adenovirus, VSV or BPV) are useful for cloning vectors in mammalian cells.

Expression and cloning vectors will typically contain a selection gene, also termed a selectable marker. Typical selection genes encode proteins that (a) confer resistance to antibiotics or other toxins, *e.g.*, ampicillin, neomycin, methotrexate, or tetracycline, (b) complement auxotrophic deficiencies, or (c) supply critical nutrients not available from complex media, *e.g.*, the gene encoding D-alanine racemase for *Bacilli*.

An example of suitable selectable markers for mammalian cells are those that enable the identification of cells competent to take up the PRO1031-encoding nucleic acid, such as DHFR or thymidine kinase. An appropriate host cell when wild-type DHFR is employed is the CHO cell line deficient in DHFR activity, prepared and propagated as described by Urlaub et al., Proc. Natl. Acad. Sci. USA, 77:4216 (1980). A suitable selection gene for use in yeast is the *trp*1 gene present in the yeast plasmid YRp7 [Stinchcomb et al., Nature, 282:39 (1979); Kingsman et al., Gene, 7:141 (1979); Tschemper et al., Gene, 10:157 (1980)]. The *trp*1 gene provides a selection marker for a mutant strain of yeast lacking the ability to grow in tryptophan, for example, ATCC No. 44076 or PEP4-1 [Jones, Genetics, 85:12 (1977)].

Expression and cloning vectors usually contain a promoter operably linked to the PRO1031-encoding nucleic acid sequence to direct mRNA synthesis. Promoters recognized by a variety of potential host cells are well known. Promoters suitable for use with prokaryotic hosts include the β-lactamase and lactose promoter systems [Chang et al., Nature, 275:615 (1978); Goeddel et al., Nature, 281:544 (1979)], alkaline phosphatase, a tryptophan (trp) promoter system [Goeddel, Nucleic Acids Res., 8:4057 (1980); EP 36,776], and hybrid promoters such as the tac promoter [deBoer et al., Proc. Natl. Acad. Sci. USA, 80:21-25 (1983)]. Promoters for use in bacterial systems also will contain a Shine-Dalgarno (S.D.) sequence operably linked to the DNA encoding the PRO1031 polypeptide.

Examples of suitable promoting sequences for use with yeast hosts include the promoters for 3-phosphoglycerate kinase [Hitzeman et al., J. Biol. Chem., 255:2073 (1980)] or other glycolytic enzymes [Hess et al., J. Adv. Enzyme Reg., 7:149 (1968); Holland, Biochemistry, 17:4900 (1978)], such as enolase, glyceraldehyde-3-phosphate dehydrogenase, hexokinase, pyruvate decarboxylase, phosphofructokinase, glucose-6-phosphate isomerase,

3-phosphoglycerate mutase, pyruvate kinase, triosephosphate isomerase, phosphoglucose isomerase, and glucokinase.

Other yeast promoters, which are inducible promoters having the additional advantage of transcription controlled by growth conditions, are the promoter regions for alcohol dehydrogenase 2, isocytochrome C, acid phosphatase, degradative enzymes associated with nitrogen metabolism, metallothionein, glyceraldehyde-3-phosphate dehydrogenase, and enzymes responsible for maltose and galactose utilization. Suitable vectors and promoters for use in yeast expression are further described in EP 73,657.

PRO1031 transcription from vectors in mammalian host cells is controlled, for example, by promoters obtained from the genomes of viruses such as polyoma virus, fowlpox virus (UK 2,211,504 published 5 July 1989), adenovirus (such as Adenovirus 2), bovine papilloma virus, avian sarcoma virus, cytomegalovirus, a retrovirus, hepatitis-B virus and Simian Virus 40 (SV40), from heterologous mammalian promoters, *e.g.*, the actin promoter or an immunoglobulin promoter, and from heat-shock promoters, provided such promoters are compatible with the host cell systems.

Transcription of a DNA encoding a PRO1031 polypeptide by higher eukaryotes may be increased by inserting an enhancer sequence into the vector. Enhancers are cis-acting elements of DNA, usually about from 10 to 300 bp, that act on a promoter to increase its transcription. Many enhancer sequences are now known from mammalian genes (globin, elastase, albumin, α-fetoprotein, and insulin). Typically, however, one will use an enhancer from a eukaryotic cell virus. Examples include the SV40 enhancer on the late side of the replication origin (bp 100-270), the cytomegalovirus early promoter enhancer, the polyoma enhancer on the late side of the replication origin, and adenovirus enhancers. The enhancer may be spliced into the vector at a position 5' or 3' to the PRO1031 coding sequence, but is preferably located at a site 5' from the promoter.

Expression vectors used in eukaryotic host cells (yeast, fungi, insect, plant, animal, human, or nucleated cells from other multicellular organisms) will also contain sequences necessary for the termination of transcription and for stabilizing the mRNA. Such sequences are commonly available from the 5' and, occasionally 3', untranslated regions of eukaryotic or viral DNAs or cDNAs. These regions contain nucleotide segments transcribed as polyadenylated fragments in the untranslated portion of the mRNA encoding PRO1031.

Still other methods, vectors, and host cells suitable for adaptation to the synthesis of PRO1031 polypeptides in recombinant vertebrate cell culture are described in Gething et al.,

Nature, 293:620-625 (1981); Mantei et al., Nature, 281:40-46 (1979); EP 117,060; and EP 117,058.

4. Detecting Gene Amplification/Expression

Gene amplification and/or expression may be measured in a sample directly, for example, by conventional Southern blotting, Northern blotting to quantitate the transcription of mRNA [Thomas, Proc. Natl. Acad. Sci. USA, 77:5201-5205 (1980)], dot blotting (DNA analysis), or *in situ* hybridization, using an appropriately labeled probe, based on the sequences provided herein. Alternatively, antibodies may be employed that can recognize specific duplexes, including DNA duplexes, RNA duplexes, and DNA-RNA hybrid duplexes or DNA-protein duplexes. The antibodies in turn may be labeled and the assay may be carried out where the duplex is bound to a surface, so that upon the formation of duplex on the surface, the presence of antibody bound to the duplex can be detected.

Gene expression, alternatively, may be measured by immunological methods, such as immunohistochemical staining of cells or tissue sections and assay of cell culture or body fluids, to quantitate directly the expression of gene product. Antibodies useful for immunohistochemical staining and/or assay of sample fluids may be either monoclonal or polyclonal, and may be prepared in any mammal. Conveniently, the antibodies may be prepared against a native sequence PRO1031 polypeptide or against a synthetic peptide based on the DNA sequences provided herein or against exogenous sequence fused to PRO1031-encoding DNA and encoding a specific antibody epitope.

5. Purification of Polypeptide

Forms of PRO1031 may be recovered from culture medium or from host cell lysates. If membrane-bound, it can be released from the membrane using a suitable detergent solution (*e.g.* Triton-X 100) or by enzymatic cleavage. Cells employed in expression of PRO1031 polypeptides can be disrupted by various physical or chemical means, such as freeze-thaw cycling, sonication, mechanical disruption, or cell lysing agents.

It may be desired to purify PRO1031 from recombinant cell proteins or polypeptides. The following procedures are exemplary of suitable purification procedures: by fractionation on an ion-exchange column; ethanol precipitation; reverse phase HPLC; chromatography on silica or on a cation-exchange resin such as DEAE; chromatofocusing; SDS-PAGE; ammonium sulfate precipitation; gel filtration using, for example, Sephadex G-75; protein

A Sepharose columns to remove contaminants such as IgG; and metal chelating columns to bind epitope-tagged forms of the PRO1031 polypeptide. Various methods of protein purification may be employed and such methods are known in the art and described for example in Deutscher, Methods in Enzymology, 182 (1990); Scopes, Protein Purification:Principles and Practice, Springer-Verlag, New York (1982). The purification step(s) selected will depend, for example, on the nature of the production process used and the particular PRO1031 polypeptide produced.

E. Uses for PRO1031

Nucleotide sequences (or their complement) encoding PRO1031 polypeptides have various applications in the art of molecular biology, including uses as hybridization probes, in chromosome and gene mapping and in the generation of anti-sense RNA and DNA. PRO1031-encoding nucleic acid will also be useful for the preparation of PRO1031 polypeptides by the recombinant techniques described herein.

The full-length DNA59294-1381 nucleotide sequence (SEQ ID NO:1) or the full-length native sequence PRO1031 nucleotide sequence, or portions thereof, may be used as hybridization probes for a cDNA library to isolate the full-length PRO1031 gene or to isolate still other genes (for instance, those encoding naturally-occurring variants of PRO1031 or PRO1031 from other species) which have a desired sequence identity to the PRO1031 nucleotide sequence disclosed in Figure 1 (SEQ ID NO:1). Optionally, the length of the probes will be about 20 to about 50 bases. The hybridization probes may be derived from the UNQ516 (DNA59294-1381) nucleotide sequence of SEQ ID NO:1 as shown in Figure 1 or from genomic sequences including promoters, enhancer elements and introns of native sequence PRO1031-encoding DNA. By way of example, a screening method will comprise isolating the coding region of the PRO1031 gene using the known DNA sequence to synthesize a selected probe of about 40 bases. Hybridization probes may be labeled by a variety of labels, including radionucleotides such as ^{32}P or ^{35}S, or enzymatic labels such as alkaline phosphatase coupled to the probe via avidin/biotin coupling systems. Labeled probes having a sequence complementary to that of the PRO1031 gene of the present invention can be used to screen libraries of human cDNA, genomic DNA or mRNA to determine which members of such libraries the probe hybridizes to. Hybridization techniques are described in further detail in the Examples below.

The probes may also be employed in PCR techniques to generate a pool of sequences for identification of closely related PRO1031 sequences.

Nucleotide sequences encoding a PRO1031 polypeptide can also be used to construct hybridization probes for mapping the gene which encodes that PRO1031 polypeptide and for the genetic analysis of individuals with genetic disorders. The nucleotide sequences provided herein may be mapped to a chromosome and specific regions of a chromosome using known techniques, such as *in situ* hybridization, linkage analysis against known chromosomal markers, and hybridization screening with libraries.

When the coding sequences for PRO1031 encode a protein which binds to another protein (example, where the PRO1031 polypeptide functions as a receptor), the PRO1031 polypeptide can be used in assays to identify the other proteins or molecules involved in the binding interaction. By such methods, inhibitors of the receptor/ligand binding interaction can be identified. Proteins involved in such binding interactions can also be used to screen for peptide or small molecule inhibitors or agonists of the binding interaction. Also, the receptor PRO1031 polypeptide can be used to isolate correlative ligand(s). Screening assays can be designed to find lead compounds that mimic the biological activity of a native PRO1031 or a receptor for PRO1031. Such screening assays will include assays amenable to high-throughput screening of chemical libraries, making them particularly suitable for identifying small molecule drug candidates. Small molecules contemplated include synthetic organic or inorganic compounds. The assays can be performed in a variety of formats, including protein-protein binding assays, biochemical screening assays, immunoassays and cell based assays, which are well characterized in the art.

Nucleic acids which encode PRO1031 polypeptide or any of its modified forms can also be used to generate either transgenic animals or "knock out" animals which, in turn, are useful in the development and screening of therapeutically useful reagents. A transgenic animal (e.g., a mouse or rat) is an animal having cells that contain a transgene, which transgene was introduced into the animal or an ancestor of the animal at a prenatal, e.g., an embryonic stage. A transgene is a DNA which is integrated into the genome of a cell from which a transgenic animal develops. In one embodiment, cDNA encoding PRO1031 polypeptide can be used to clone genomic DNA encoding PRO1031 in accordance with established techniques and the genomic sequences used to generate transgenic animals that contain cells which express DNA encoding PRO1031. Methods for generating transgenic animals, particularly animals such as mice or rats, have become conventional in the art and

are described, for example, in U.S. Patent Nos. 4,736,866 and 4,870,009. Typically, particular cells would be targeted for PRO1031 transgene incorporation with tissue-specific enhancers. Transgenic animals that include a copy of a transgene encoding PRO1031 introduced into the germ line of the animal at an embryonic stage can be used to examine the effect of increased expression of DNA encoding PRO1031. Such animals can be used as tester animals for reagents thought to confer protection from, for example, pathological conditions associated with its overexpression. In accordance with this facet of the invention, an animal is treated with the reagent and a reduced incidence of the pathological condition, compared to untreated animals bearing the transgene, would indicate a potential therapeutic intervention for the pathological condition.

Alternatively, non-human homologues of PRO1031 can be used to construct a PRO1031 "knock out" animal which has a defective or altered gene encoding PRO1031 as a result of homologous recombination between the endogenous gene encoding PRO1031 and altered genomic DNA encoding PRO1031 introduced into an embryonic cell of the animal. For example, cDNA encoding PRO1031 can be used to clone genomic DNA encoding PRO1031 in accordance with established techniques. A portion of the genomic DNA encoding PRO1031 can be deleted or replaced with another gene, such as a gene encoding a selectable marker which can be used to monitor integration. Typically, several kilobases of unaltered flanking DNA (both at the 5' and 3' ends) are included in the vector [see e.g., Thomas and Capecchi, Cell, 51:503 (1987) for a description of homologous recombination vectors]. The vector is introduced into an embryonic stem cell line (e.g., by electroporation) and cells in which the introduced DNA has homologously recombined with the endogenous DNA are selected [see e.g., Li et al., Cell, 69:915 (1992)]. The selected cells are then injected into a blastocyst of an animal (e.g., a mouse or rat) to form aggregation chimeras [see e.g., Bradley, in *Teratocarcinomas and Embryonic Stem Cells: A Practical Approach*, E. J. Robertson, ed. (IRL, Oxford, 1987), pp. 113-152]. A chimeric embryo can then be implanted into a suitable pseudopregnant female foster animal and the embryo brought to term to create a "knock out" animal. Progeny harboring the homologously recombined DNA in their germ cells can be identified by standard techniques and used to breed animals in which all cells of the animal contain the homologously recombined DNA. Knockout animals can be characterized for instance, for their ability to defend against certain pathological conditions and for their development of pathological conditions due to absence of the PRO1031 polypeptide.

PRO1031 polypeptides of the present invention which possess biological activity related to that of IL-17 may be employed both *in vivo* for therapeutic purposes and *in vitro*. Those of ordinary skill in the art will well know how to employ the PRO1031 polypeptides of the present invention for such purposes.

PRO1031 can be used in assays with the polypeptides to which they have identity with to determine the relative activities. The results can be applied accordingly.

F. Anti-PRO1031 Antibodies

The present invention further provides anti-PRO1031 polypeptide antibodies. Exemplary antibodies include polyclonal, monoclonal, humanized, bispecific, and heteroconjugate antibodies.

1. Polyclonal Antibodies

The anti-PRO1031 antibodies of the present invention may comprise polyclonal antibodies. Methods of preparing polyclonal antibodies are known to the skilled artisan. Polyclonal antibodies can be raised in a mammal, for example, by one or more injections of an immunizing agent and, if desired, an adjuvant. Typically, the immunizing agent and/or adjuvant will be injected in the mammal by multiple subcutaneous or intraperitoneal injections. The immunizing agent may include the PRO1031 polypeptide or a fusion protein thereof. It may be useful to conjugate the immunizing agent to a protein known to be immunogenic in the mammal being immunized. Examples of such immunogenic proteins include but are not limited to keyhole limpet hemocyanin, serum albumin, bovine thyroglobulin, and soybean trypsin inhibitor. Examples of adjuvants which may be employed include Freund's complete adjuvant and MPL-TDM adjuvant (monophosphoryl Lipid A, synthetic trehalose dicorynomycolate). The immunization protocol may be selected by one skilled in the art without undue experimentation.

2. Monoclonal Antibodies

The anti-PRO1031 antibodies may, alternatively, be monoclonal antibodies. Monoclonal antibodies may be prepared using hybridoma methods, such as those described by Kohler and Milstein, Nature, 256:495 (1975). In a hybridoma method, a mouse, hamster, or other appropriate host animal, is typically immunized with an immunizing agent to elicit

lymphocytes that produce or are capable of producing antibodies that will specifically bind to the immunizing agent. Alternatively, the lymphocytes may be immunized *in vitro*.

The immunizing agent will typically include the PRO1031 polypeptide or a fusion protein thereof. Generally, either peripheral blood lymphocytes ("PBLs") are used if cells of human origin are desired, or spleen cells or lymph node cells are used if non-human mammalian sources are desired. The lymphocytes are then fused with an immortalized cell line using a suitable fusing agent, such as polyethylene glycol, to form a hybridoma cell [Goding, Monoclonal Antibodies: Principles and Practice, Academic Press, (1986) pp. 59-103]. Immortalized cell lines are usually transformed mammalian cells, particularly myeloma cells of rodent, bovine and human origin. Usually, rat or mouse myeloma cell lines are employed. The hybridoma cells may be cultured in a suitable culture medium that preferably contains one or more substances that inhibit the growth or survival of the unfused, immortalized cells. For example, if the parental cells lack the enzyme hypoxanthine guanine phosphoribosyl transferase (HGPRT or HPRT), the culture medium for the hybridomas typically will include hypoxanthine, aminopterin, and thymidine ("HAT medium"), which substances prevent the growth of HGPRT-deficient cells.

Preferred immortalized cell lines are those that fuse efficiently, support stable high level expression of antibody by the selected antibody-producing cells, and are sensitive to a medium such as HAT medium. More preferred immortalized cell lines are murine myeloma lines, which can be obtained, for instance, from the Salk Institute Cell Distribution Center, San Diego, California and the American Type Culture Collection, Rockville, Maryland. Human myeloma and mouse-human heteromyeloma cell lines also have been described for the production of human monoclonal antibodies [Kozbor, J. Immunol., 133:3001 (1984); Brodeur et al., Monoclonal Antibody Production Techniques and Applications, Marcel Dekker, Inc., New York, (1987) pp. 51-63].

The culture medium in which the hybridoma cells are cultured can then be assayed for the presence of monoclonal antibodies directed against a PRO1031 polypeptide. Preferably, the binding specificity of monoclonal antibodies produced by the hybridoma cells is determined by immunoprecipitation or by an *in vitro* binding assay, such as radioimmunoassay (RIA) or enzyme-linked immunoabsorbent assay (ELISA). Such techniques and assays are known in the art. The binding affinity of the monoclonal antibody can, for example, be determined by the Scatchard analysis of Munson and Pollard, Anal. Biochem., 107:220 (1980).

After the desired hybridoma cells are identified, the clones may be subcloned by limiting dilution procedures and grown by standard methods [Goding, supra]. Suitable culture media for this purpose include, for example, Dulbecco's Modified Eagle's Medium and RPMI-1640 medium. Alternatively, the hybridoma cells may be grown *in vivo* as ascites in a mammal.

The monoclonal antibodies secreted by the subclones may be isolated or purified from the culture medium or ascites fluid by conventional immunoglobulin purification procedures such as, for example, protein A-Sepharose, hydroxylapatite chromatography, gel electrophoresis, dialysis, or affinity chromatography.

The monoclonal antibodies may also be made by recombinant DNA methods, such as those described in U.S. Patent No. 4,816,567. DNA encoding the monoclonal antibodies of the invention can be readily isolated and sequenced using conventional procedures (e.g., by using oligonucleotide probes that are capable of binding specifically to genes encoding the heavy and light chains of murine antibodies). The hybridoma cells of the invention serve as a preferred source of such DNA. Once isolated, the DNA may be placed into expression vectors, which are then transfected into host cells such as simian COS cells, Chinese hamster ovary (CHO) cells, or myeloma cells that do not otherwise produce immunoglobulin protein, to obtain the synthesis of monoclonal antibodies in the recombinant host cells. The DNA also may be modified, for example, by substituting the coding sequence for human heavy and light chain constant domains in place of the homologous murine sequences [U.S. Patent No. 4,816,567; Morrison et al., supra] or by covalently joining to the immunoglobulin coding sequence all or part of the coding sequence for a non-immunoglobulin polypeptide. Such a non-immunoglobulin polypeptide can be substituted for the constant domains of an antibody of the invention, or can be substituted for the variable domains of one antigen-combining site of an antibody of the invention to create a chimeric bivalent antibody.

The antibodies may be monovalent antibodies. Methods for preparing monovalent antibodies are well known in the art. For example, one method involves recombinant expression of immunoglobulin light chain and modified heavy chain. The heavy chain is truncated generally at any point in the Fc region so as to prevent heavy chain crosslinking. Alternatively, the relevant cysteine residues are substituted with another amino acid residue or are deleted so as to prevent crosslinking.

In vitro methods are also suitable for preparing monovalent antibodies. Digestion of antibodies to produce fragments thereof, particularly, Fab fragments, can be accomplished using routine techniques known in the art.

3. Humanized Antibodies

The anti-PRO1031 antibodies of the invention may further comprise humanized antibodies or human antibodies. Humanized forms of non-human (e.g., murine) antibodies are chimeric immunoglobulins, immunoglobulin chains or fragments thereof (such as Fv, Fab, Fab', $F(ab')_2$ or other antigen-binding subsequences of antibodies) which contain minimal sequence derived from non-human immunoglobulin. Humanized antibodies include human immunoglobulins (recipient antibody) in which residues from a complementary determining region (CDR) of the recipient are replaced by residues from a CDR of a non-human species (donor antibody) such as mouse, rat or rabbit having the desired specificity, affinity and capacity. In some instances, Fv framework residues of the human immunoglobulin are replaced by corresponding non-human residues. Humanized antibodies may also comprise residues which are found neither in the recipient antibody nor in the imported CDR or framework sequences. In general, the humanized antibody will comprise substantially all of at least one, and typically two, variable domains, in which all or substantially all of the CDR regions correspond to those of a non-human immunoglobulin and all or substantially all of the FR regions are those of a human immunoglobulin consensus sequence. The humanized antibody optimally also will comprise at least a portion of an immunoglobulin constant region (Fc), typically that of a human immunoglobulin [Jones et al., Nature, 321:522-525 (1986); Riechmann et al., Nature, 332:323-329 (1988); and Presta, Curr. Op. Struct. Biol., 2:593-596 (1992)].

Methods for humanizing non-human antibodies are well known in the art. Generally, a humanized antibody has one or more amino acid residues introduced into it from a source which is non-human. These non-human amino acid residues are often referred to as "import" residues, which are typically taken from an "import" variable domain. Humanization can be essentially performed following the method of Winter and co-workers [Jones et al., Nature, 321:522-525 (1986); Riechmann et al., Nature, 332:323-327 (1988); Verhoeyen et al., Science, 239:1534-1536 (1988)], by substituting rodent CDRs or CDR sequences for the corresponding sequences of a human antibody. Accordingly, such "humanized" antibodies are chimeric antibodies (U.S. Patent No. 4,816,567), wherein substantially less than an intact

human variable domain has been substituted by the corresponding sequence from a non-human species. In practice, humanized antibodies are typically human antibodies in which some CDR residues and possibly some FR residues are substituted by residues from analogous sites in rodent antibodies.

Human antibodies can also be produced using various techniques known in the art, including phage display libraries [Hoogenboom and Winter, J. Mol. Biol., 227:381 (1991); Marks et al., J. Mol. Biol., 222:581 (1991)]. The techniques of Cole et al. and Boerner et al. are also available for the preparation of human monoclonal antibodies (Cole et al., Monoclonal Antibodies and Cancer Therapy, Alan R. Liss, p. 77 (1985) and Boerner et al., J. Immunol., 147(1):86-95 (1991)].

4. Bispecific Antibodies

Bispecific antibodies are monoclonal, preferably human or humanized, antibodies that have binding specificities for at least two different antigens. In the present case, one of the binding specificities is for a PRO1031 polypeptide, the other one is for any other antigen, and preferably for a cell-surface protein or receptor or receptor subunit.

Methods for making bispecific antibodies are known in the art. Traditionally, the recombinant production of bispecific antibodies is based on the co-expression of two immunoglobulin heavy-chain/light-chain pairs, where the two heavy chains have different specificities [Milstein and Cuello, Nature, 305:537-539 (1983)]. Because of the random assortment of immunoglobulin heavy and light chains, these hybridomas (quadromas) produce a potential mixture of ten different antibody molecules, of which only one has the correct bispecific structure. The purification of the correct molecule is usually accomplished by affinity chromatography steps. Similar procedures are disclosed in WO 93/08829, published 13 May 1993, and in Traunecker et al., EMBO J., 10:3655-3659 (1991).

Antibody variable domains with the desired binding specificities (antibody-antigen combining sites) can be fused to immunoglobulin constant domain sequences. The fusion preferably is with an immunoglobulin heavy-chain constant domain, comprising at least part of the hinge, CH2, and CH3 regions. It is preferred to have the first heavy-chain constant region (CH1) containing the site necessary for light-chain binding present in at least one of the fusions. DNAs encoding the immunoglobulin heavy-chain fusions and, if desired, the immunoglobulin light chain, are inserted into separate expression vectors, and are co-

transfected into a suitable host organism. For further details of generating bispecific antibodies see, for example, Suresh et al., Methods in Enzymology, 121:210 (1986).

5. Heteroconjugate Antibodies

Heteroconjugate antibodies are also within the scope of the present invention. Heteroconjugate antibodies are composed of two covalently joined antibodies. Such antibodies have, for example, been proposed to target immune system cells to unwanted cells [U.S. Patent No. 4,676,980], and for treatment of HIV infection [WO 91/00360; WO 92/200373; EP 03089]. It is contemplated that the antibodies may be prepared *in vitro* using known methods in synthetic protein chemistry, including those involving crosslinking agents. For example, immunotoxins may be constructed using a disulfide exchange reaction or by forming a thioether bond. Examples of suitable reagents for this purpose include iminothiolate and methyl-4-mercaptobutyrimidate and those disclosed, for example, in U.S. Patent No. 4,676,980.

G. Uses for anti-PRO1031 Antibodies

The anti-PRO1031 antibodies of the present invention have various utilities. For example, anti-PRO1031 antibodies may be used in diagnostic assays for PRO1031 polypeptides, *e.g.*, detecting expression in specific cells, tissues, or serum. Various diagnostic assay techniques known in the art may be used, such as competitive binding assays, direct or indirect sandwich assays and immunoprecipitation assays conducted in either heterogeneous or homogeneous phases [Zola, Monoclonal Antibodies: A Manual of Techniques, CRC Press, Inc. (1987) pp. 147-158]. The antibodies used in the diagnostic assays can be labeled with a detectable moiety. The detectable moiety should be capable of producing, either directly or indirectly, a detectable signal. For example, the detectable moiety may be a radioisotope, such as ^{3}H, ^{14}C, ^{32}P, ^{35}S, or ^{125}I, a fluorescent or chemiluminescent compound, such as fluorescein isothiocyanate, rhodamine, or luciferin, or an enzyme, such as alkaline phosphatase, beta-galactosidase or horseradish peroxidase. Any method known in the art for conjugating the antibody to the detectable moiety may be employed, including those methods described by Hunter et al., Nature, 144:945 (1962); David et al., Biochemistry, 13:1014 (1974); Pain et al., J. Immunol. Meth., 40:219 (1981); and Nygren, J. Histochem. and Cytochem., 30:407 (1982).

Anti-PRO1031 antibodies also are useful for the affinity purification of PRO1031 polypeptides from recombinant cell culture or natural sources. In this process, the antibodies against a PRO1031 polypeptide are immobilized on a suitable support, such a Sephadex resin or filter paper, using methods well known in the art. The immobilized antibody then is contacted with a sample containing the PRO1031 polypeptide to be purified, and thereafter the support is washed with a suitable solvent that will remove substantially all the material in the sample except the PRO1031 polypeptide, which is bound to the immobilized antibody. Finally, the support is washed with another suitable solvent that will release the PRO1031 polypeptide from the antibody.

The following examples are offered for illustrative purposes only, and are not intended to limit the scope of the present invention in any way.

All patent and literature references cited in the present specification are hereby incorporated by reference in their entirety.

EXAMPLES

Commercially available reagents referred to in the examples were used according to manufacturer's instructions unless otherwise indicated. The source of those cells identified in the following examples, and throughout the specification, by ATCC accession numbers is the American Type Culture Collection, Rockville, Maryland.

EXAMPLE 1: Isolation of cDNA Clones Encoding Human PRO1031

The extracellular domain (ECD) sequences (including the secretion signal, if any) of from about 950 known secreted proteins from the Swiss-Prot public protein database were used to search expressed sequence tag (EST) databases. The EST databases included public EST databases (e.g., GenBank, Merck/Wash U.) and a proprietary EST DNA database (LIFESEQ™, Incyte Pharmaceuticals, Palo Alto, CA). The search was performed using the computer program BLAST or BLAST2 (Altshul et al., Methods in Enzymology 266:460-480 (1996)) as a comparison of the ECD protein sequences to a 6 frame translation of the EST sequence. Those comparisons resulting in a BLAST score of 70 (or in some cases 90) or greater that did not encode known proteins were clustered and assembled into consensus DNA sequences with the program "phrap" (Phil Green, University of Washington, Seattle, Washington; http://bozeman.mbt.washington.edu/phrap.docs/phrap.html).

An initial consensus DNA sequence was assembled relative to other EST sequences using phrap. The initial consensus DNA sequence was extended using repeated cycles of BLAST and phrap to extend the consensus sequence as far as possible using the sources of EST sequences discussed above. Figure 3 shows the assembly, wherein the initial consensus sequence is shown as "DNA47332.init" and the extended consensus sequence is shown as "<consen01>" (also sometimes referred to as DNA47332).

Based on the consensus sequences assembled herein and other information provided herein, Merck W74558 (clone 344649) was further examined. DNA sequencing gave the full-length DNA sequence for PRO1031 [herein designated as UNQ516 (DNA59294-1381)] (SEQ ID NO:1) and the derived protein sequence for PRO1031.

The entire nucleotide sequence of UNQ516 (DNA59294-1381) is shown in Figure 1 (SEQ ID NO:1). Clone UNQ516 (DNA59294-1381) contains a single open reading frame with an apparent translational initiation site at nucleotide positions 42-44 [Kozak et al., supra] and ending at the stop codon at nucleotide positions 582-584 (Figure 1). The predicted polypeptide precursor is 180 amino acids long (Figure 2). The full-length PRO1031 protein shown in Figure 2 has an estimated molecular weight of about 20437 and a pI of about 9.58. Clone UNQ516 (DNA59294-1381) has been deposited with the ATCC. Regarding the sequence, it is understood that the deposited clone contains the correct sequence, and the sequences provided herein are based on known sequencing techniques.

Analysis of the amino acid sequence of the full-length PRO1031 polypeptide suggests that it is a novel cytokine.

Still analyzing the amino acid sequence of SEQ ID NO:2, the putative signal peptide is at about amino acids 1-20 of SEQ ID NO:2. An N-glycosylation site is at about amino acids 75-78 of SEQ ID NO:2. A region having sequence identity with IL-17 is at about amino acids 96-180. The corresponding nucleotides can be routinely determined given the sequences provided herein.

EXAMPLE 2: Use of PRO1031-encoding DNA as a hybridization probe

The following method describes use of a nucleotide sequence encoding PRO1031 as a hybridization probe.

DNA comprising the coding sequence of full-length PRO1031 (as shown in Figure 1, SEQ ID NO:1) or a fragment thereof is employed as a probe to screen for homologous

DNAs (such as those encoding naturally-occurring variants of PRO1031) in human tissue cDNA libraries or human tissue genomic libraries.

Hybridization and washing of filters containing either library DNAs is performed under the following high stringency conditions. Hybridization of radiolabeled PRO1031 polypeptide-derived probe to the filters is performed in a solution of 50% formamide, 5 x SSC, 0.1% SDS, 0.1% sodium pyrophosphate, 50 mM sodium phosphate, pH 6.8, 2 x Denhardt's solution, and 10% dextran sulfate at 42°C for 20 hours. Washing of the filters is performed in an aqueous solution of 0.1 x SSC and 0.1% SDS at 42°C.

DNAs having a desired sequence identity with the DNA encoding full-length native sequence PRO1031 polypeptide can then be identified using standard techniques known in the art.

EXAMPLE 3: Expression of PRO1031 Polypeptides in *E. coli*

This example illustrates the preparation of unglycosylated forms of PRO1031 polypeptides by recombinant expression in *E. coli*.

The DNA sequence encoding the full-length PRO1031 or a fragment or variant thereof is initially amplified using selected PCR primers. The primers should contain restriction enzyme sites which correspond to the restriction enzyme sites on the selected expression vector. A variety of expression vectors may be employed. An example of a suitable vector is pBR322 (derived from *E. coli*; see Bolivar et al., Gene, 2:95 (1977)) which contains genes for ampicillin and tetracycline resistance. The vector is digested with restriction enzyme and dephosphorylated. The PCR amplified sequences are then ligated into the vector. The vector will preferably include sequences which encode for an antibiotic resistance gene, a trp promoter, a polyhis leader (including the first six STII codons, polyhis sequence, and enterokinase cleavage site), the PRO1031 coding region, lambda transcriptional terminator, and an argU gene.

The ligation mixture is then used to transform a selected *E. coli* strain using the methods described in Sambrook et al., supra. Transformants are identified by their ability to grow on LB plates and antibiotic resistant colonies are then selected. Plasmid DNA can be isolated and confirmed by restriction analysis and DNA sequencing.

Selected clones can be grown overnight in liquid culture medium such as LB broth supplemented with antibiotics. The overnight culture may subsequently be used to inoculate

a larger scale culture. The cells are then grown to a desired optical density, during which the expression promoter is turned on.

After culturing the cells for several more hours, the cells can be harvested by centrifugation. The cell pellet obtained by the centrifugation can be solubilized using various agents known in the art, and the solubilized PRO1031 polypeptide can then be purified using a metal chelating column under conditions that allow tight binding of the polypeptide.

EXAMPLE 4: Expression of PRO1031 Polypeptides in Mammalian Cells

This example illustrates preparation of glycosylated forms of PRO1031 polypeptides by recombinant expression in mammalian cells.

The vector, pRK5 (see EP 307,247, published March 15, 1989), is employed as the expression vector. Optionally, the PRO1031-encoding DNA is ligated into pRK5 with selected restriction enzymes to allow insertion of the PRO1031-encoding DNA using ligation methods such as described in Sambrook et al., supra. The resulting vector is called pRK5-PRO1031.

In one embodiment, the selected host cells may be 293 cells. Human 293 cells (ATCC CCL 1573) are grown to confluence in tissue culture plates in medium such as DMEM supplemented with fetal calf serum and optionally, nutrient components and/or antibiotics. About 10 μg pRK5-PRO1031 DNA is mixed with about 1 μg DNA encoding the VA RNA gene [Thimmappaya et al., Cell, 31:543 (1982)] and dissolved in 500 μl of 1 mM Tris-HCl, 0.1 mM EDTA, 0.227 M $CaCl_2$. To this mixture is added, dropwise, 500 μl of 50 mM HEPES (pH 7.35), 280 mM NaCl, 1.5 mM $NaPO_4$, and a precipitate is allowed to form for 10 minutes at 25°C. The precipitate is suspended and added to the 293 cells and allowed to settle for about four hours at 37°C. The culture medium is aspirated off and 2 ml of 20% glycerol in PBS is added for 30 seconds. The 293 cells are then washed with serum free medium, fresh medium is added and the cells are incubated for about 5 days.

Approximately 24 hours after the transfections, the culture medium is removed and replaced with culture medium (alone) or culture medium containing 200 μCi/ml ^{35}S-cysteine and 200 μCi/ml ^{35}S-methionine. After a 12 hour incubation, the conditioned medium is collected, concentrated on a spin filter, and loaded onto a 15% SDS gel. The processed gel may be dried and exposed to film for a selected period of time to reveal the presence of PRO1031 polypeptide. The cultures containing transfected cells may undergo further incubation (in serum free medium) and the medium is tested in selected bioassays.

In an alternative technique, PRO1031-encoding DNA may be introduced into 293 cells transiently using the dextran sulfate method described by Somparyrac et al., Proc. Natl. Acad. Sci., 12:7575 (1981). 293 cells are grown to maximal density in a spinner flask and 700 μg pRK5-PRO1031 DNA is added. The cells are first concentrated from the spinner flask by centrifugation and washed with PBS. The DNA-dextran precipitate is incubated on the cell pellet for four hours. The cells are treated with 20% glycerol for 90 seconds, washed with tissue culture medium, and re-introduced into the spinner flask containing tissue culture medium, 5 μg/ml bovine insulin and 0.1 μg/ml bovine transferrin. After about four days, the conditioned media is centrifuged and filtered to remove cells and debris. The sample containing expressed PRO1031 polypeptide can then be concentrated and purified by any selected method, such as dialysis and/or column chromatography.

In another embodiment, PRO1031 polypeptide can be expressed in CHO cells. The pRK5-PRO1031 vector can be transfected into CHO cells using known reagents such as $CaPO_4$ or DEAE-dextran. As described above, the cell cultures can be incubated, and the medium replaced with culture medium (alone) or medium containing a radiolabel such as ^{35}S-methionine. After determining the presence of PRO1031 polypeptide, the culture medium may be replaced with serum free medium. Preferably, the cultures are incubated for about 6 days, and then the conditioned medium is harvested. The medium containing the expressed PRO1031 polypeptide can then be concentrated and purified by any selected method.

Epitope-tagged PRO1031 polypeptide may also be expressed in host CHO cells. The PRO1031-encoding DNA may be subcloned out of the pRK5 vector. The subclone insert can undergo PCR to fuse in frame with a selected epitope tag such as a poly-his tag into a Baculovirus expression vector. The poly-his tagged PRO1031-encoding DNA insert can then be subcloned into a SV40 driven vector containing a selection marker such as DHFR for selection of stable clones. Finally, the CHO cells can be transfected (as described above) with the SV40 driven vector. Labeling may be performed, as described above, to verify expression. The culture medium containing the expressed poly-His tagged PRO1031 polypeptide can then be concentrated and purified by any selected method, such as by Ni^{2+}-chelate affinity chromatography.

EXAMPLE 5: Expression of a PRO1031 Polypeptide in Yeast

The following method describes recombinant expression of PRO1031 polypeptides in yeast.

First, yeast expression vectors are constructed for intracellular production or secretion of PRO1031 polypeptide from the ADH2/GAPDH promoter. DNA encoding the PRO1031 polypeptide of interest, a selected signal peptide and the promoter is inserted into suitable restriction enzyme sites in the selected plasmid to direct intracellular expression of the PRO1031 polypeptide. For secretion, DNA encoding the PRO1031 polypeptide can be cloned into the selected plasmid, together with DNA encoding the ADH2/GAPDH promoter, the yeast alpha-factor secretory signal/leader sequence, and linker sequences (if needed) for expression of the PRO1031 polypeptide.

Yeast cells, such as yeast strain AB110, can then be transformed with the expression plasmids described above and cultured in selected fermentation media. The transformed yeast supernatants can be analyzed by precipitation with 10% trichloroacetic acid and separation by SDS-PAGE, followed by staining of the gels with Coomassie Blue stain.

Recombinant PRO1031 polypeptide can subsequently be isolated and purified by removing the yeast cells from the fermentation medium by centrifugation and then concentrating the medium using selected cartridge filters. The concentrate containing the PRO1031 polypeptide may further be purified using selected column chromatography resins.

EXAMPLE 6: Expression of PRO1031 Polypeptides in Baculovirus-Infected Insect Cells

The following method describes recombinant expression of PRO1031 polypeptides in Baculovirus-infected insect cells.

The PRO1031-encoding DNA is fused upstream of an epitope tag contained within a baculovirus expression vector. Such epitope tags include poly-his tags and immunoglobulin tags (like Fc regions of IgG). A variety of plasmids may be employed, including plasmids derived from commercially available plasmids such as pVL1393 (Novagen). Briefly, the PRO1031-encoding DNA or the desired portion of the PRO1031-encoding DNA (such as the sequence encoding the extracellular domain of a transmembrane protein) is amplified by PCR with primers complementary to the 5' and 3' regions. The 5' primer may incorporate flanking (selected) restriction enzyme sites. The product is then digested with those selected restriction enzymes and subcloned into the expression vector.

Recombinant baculovirus is generated by co-transfecting the above plasmid and BaculoGold™ virus DNA (Pharmingen) into *Spodoptera frugiperda* ("Sf9") cells (ATCC CRL 1711) using lipofectin (commercially available from GIBCO-BRL). After 4 to 5 days of incubation at 28°C, the released viruses are harvested and used for further amplifications.

Viral infection and protein expression is performed as described by O'Reilley et al., Baculovirus expression vectors: A laboratory Manual, Oxford:Oxford University Press (1994).

Expressed poly-his tagged PRO1031 polypeptide can then be purified, for example, by Ni^{2+}-chelate affinity chromatography as follows. Extracts are prepared from recombinant virus-infected Sf9 cells as described by Rupert et al., Nature, 362:175-179 (1993). Briefly, Sf9 cells are washed, resuspended in sonication buffer (25 mL Hepes, pH 7.9; 12.5 mM $MgCl_2$; 0.1 mM EDTA; 10% Glycerol; 0.1% NP-40; 0.4 M KCl), and sonicated twice for 20 seconds on ice. The sonicates are cleared by centrifugation, and the supernatant is diluted 50-fold in loading buffer (50 mM phosphate, 300 mM NaCl, 10% Glycerol, pH 7.8) and filtered through a 0.45 μm filter. A Ni^{2+}-NTA agarose column (commercially available from Qiagen) is prepared with a bed volume of 5 mL, washed with 25 mL of water and equilibrated with 25 mL of loading buffer. The filtered cell extract is loaded onto the column at 0.5 mL per minute. The column is washed to baseline A_{280} with loading buffer, at which point fraction collection is started. Next, the column is washed with a secondary wash buffer (50 mM phosphate; 300 mM NaCl, 10% Glycerol, pH 6.0), which elutes nonspecifically bound protein. After reaching A_{280} baseline again, the column is developed with a 0 to 500 mM Imidazole gradient in the secondary wash buffer. One mL fractions are collected and analyzed by SDS-PAGE and silver staining or western blot with Ni^{2+}-NTA-conjugated to alkaline phosphatase (Qiagen). Fractions containing the eluted His_{10}-tagged PRO1031 polypeptide are pooled and dialyzed against loading buffer.

Alternatively, purification of the IgG tagged (or Fc tagged) PRO1031 polypeptide can be performed using known chromatography techniques, including for instance, Protein A or protein G column chromatography.

EXAMPLE 7: Preparation of Antibodies that Bind PRO1031 Polypeptides

This example illustrates the preparation of monoclonal antibodies which can specifically bind to PRO1031 polypeptides.

Techniques for producing the monoclonal antibodies are known in the art and are described, for instance, in Goding, supra. Immunogens that may be employed include purified PRO1031 polypeptide, fusion proteins containing a PRO1031 polypeptide, and cells expressing recombinant PRO1031 polypeptide on the cell surface. Selection of the immunogen can be made by the skilled artisan without undue experimentation.

Mice, such as Balb/c, are immunized with the PRO1031 immunogen emulsified in complete Freund's adjuvant and injected subcutaneously or intraperitoneally in an amount from 1-100 micrograms. Alternatively, the immunogen is emulsified in MPL-TDM adjuvant (Ribi Immunochemical Research, Hamilton, MT) and injected into the animal's hind foot pads. The immunized mice are then boosted 10 to 12 days later with additional immunogen emulsified in the selected adjuvant. Thereafter, for several weeks, the mice may also be boosted with additional immunization injections. Serum samples may be periodically obtained from the mice by retro-orbital bleeding for testing in ELISA assays to detect anti-PRO1031 polypeptide antibodies.

After a suitable antibody titer has been detected, the animals "positive" for antibodies can be injected with a final intravenous injection of PRO1031 polypeptide. Three to four days later, the mice are sacrificed and the spleen cells are harvested. The spleen cells are then fused (using 35% polyethylene glycol) to a selected murine myeloma cell line such as P3X63AgU.1, available from ATCC, No. CRL 1597. The fusions generate hybridoma cells which can then be plated in 96 well tissue culture plates containing HAT (hypoxanthine, aminopterin, and thymidine) medium to inhibit proliferation of non-fused cells, myeloma hybrids, and spleen cell hybrids.

The hybridoma cells will be screened in an ELISA for reactivity against PRO1031 polypeptide. Determination of "positive" hybridoma cells secreting the desired monoclonal antibodies against a PRO1031 polypeptide is within the skill in the art.

The positive hybridoma cells can be injected intraperitoneally into syngeneic Balb/c mice to produce ascites containing the anti-PRO1031 polypeptide monoclonal antibodies. Alternatively, the hybridoma cells can be grown in tissue culture flasks or roller bottles. Purification of the monoclonal antibodies produced in the ascites can be accomplished using ammonium sulfate precipitation, followed by gel exclusion chromatography. Alternatively, affinity chromatography based upon binding of antibody to protein A or protein G can be employed.

Deposit of Material

The following materials have been deposited with the American Type Culture Collection, 12301 Parklawn Drive, Rockville, MD, USA (ATCC):

Material	ATCC Dep. No.	Deposit Date
DNA59294-1381		14 May 1998

This deposit was made under the provisions of the Budapest Treaty on the International Recognition of the Deposit of Microorganisms for the Purpose of Patent Procedure and the Regulations thereunder (Budapest Treaty). This assures maintenance of a viable culture of the deposit for 30 years from the date of deposit. The deposit will be made available by ATCC under the terms of the Budapest Treaty, and subject to an agreement between Genentech, Inc. and ATCC, which assures permanent and unrestricted availability of the progeny of the culture of the deposit to the public upon issuance of the pertinent U.S. patent or upon laying open to the public of any U.S. or foreign patent application, whichever comes first, and assures availability of the progeny to one determined by the U.S. Commissioner of Patents and Trademarks to be entitled thereto according to 35 USC §122 and the Commissioner's rules pursuant thereto (including 37 CFR §1.14 with particular reference to 886 OG 638).

The assignee of the present application has agreed that if a culture of the materials on deposit should die or be lost or destroyed when cultivated under suitable conditions, the materials will be promptly replaced on notification with another of the same. Availability of the deposited material is not to be construed as a license to practice the invention in contravention of the rights granted under the authority of any government in accordance with its patent laws.

The foregoing written specification is considered to be sufficient to enable one skilled in the art to practice the invention. The present invention is not to be limited in scope by the construct deposited, since the deposited embodiment is intended as a single illustration of certain aspects of the invention and any constructs that are functionally equivalent are within the scope of this invention. The deposit of material herein does not constitute an admission that the written description herein contained is inadequate to enable the practice of any aspect of the invention, including the best mode thereof, nor is it to be construed as limiting the scope of the claims to the specific illustrations that it represents. Indeed, various modifications of the invention in addition to those shown and described herein will become apparent to those skilled in the art from the foregoing description and fall within the scope of the appended claims.

WHAT IS CLAIMED IS:

1. Isolated nucleic acid comprising DNA having at least an 80% sequence identity to (a) a DNA molecule encoding a PRO1031 polypeptide comprising the sequence of amino acid residues 1 or 21 through 180 of Figure 2 (SEQ ID NO:2), or (b) the complement of the DNA molecule of (a).

2. The nucleic acid of Claim 1, wherein said DNA comprises nucleotides 42 or 102 through 581 of SEQ ID NO:1.

3. The nucleic acid of Claim 1, wherein said DNA comprises the nucleotide sequence of SEQ ID NO:1.

4. Isolated nucleic acid comprising DNA having at least an 80% sequence identity to (a) a DNA molecule encoding the same mature polypeptide encoded by the human protein cDNA in ATCC Deposit of DNA59294-1381, or (b) the complement of the DNA molecule of (a).

5. The nucleic acid of Claim 4 which comprises a DNA molecule encoding the same mature polypeptide encoded by the human protein cDNA in said ATCC Deposit.

6. A vector comprising the nucleic acid of any one of Claims 1 through 5.

7. The vector of Claim 6 operably linked to control sequences recognized by a host cell transformed with the vector.

8. A host cell comprising the vector of Claim 7.

9. The host cell of Claim 8, wherein said cell is a CHO cell.

10. The host cell of Claim 8, wherein said cell is an *E. coli*.

11. The host cell of Claim 8, wherein said cell is a yeast cell.

12. A process for producing a PRO1031 polypeptide comprising culturing the host cell of Claim 8 under conditions suitable for expression of said PRO1031 polypeptide and recovering said PRO1031 polypeptide from the cell culture.

13. An isolated polypeptide comprising amino acid residues 1 or 21 through 180 of Figure 2 (SEQ ID NO:2).

14. Isolated PRO1031 polypeptide encoded by the cDNA insert of the vector deposited with the ATCC as DNA59294-1381.

15. A chimeric molecule comprising a PRO1031 polypeptide fused to a heterologous amino acid sequence.

16. The chimeric molecule of Claim 15, wherein said heterologous amino acid sequence is an epitope tag sequence.

17. The chimeric molecule of Claim 15, wherein said heterologous amino acid sequence is a Fc region of an immunoglobulin.

18. An antibody which specifically binds to a PRO1031 polypeptide.

19. The antibody of Claim 18, wherein said antibody is a monoclonal antibody.

Abstract of the Disclosure

The present invention is directed to novel polypeptides having sequence identity with IL-17 and to nucleic acid molecules encoding those polypeptides. Also provided herein are vectors and host cells comprising those nucleic acid sequences, chimeric polypeptide molecules comprising the polypeptides of the present invention fused to heterologous polypeptide sequences, antibodies which bind to the polypeptides of the present invention and to methods for producing the polypeptides of the present invention.

SEQ. ID NO:1

```
   1 AGGCGGGCAG CAGCTGCAGG CTGACCTTGC AGCTTGGCGG AATGGACTGG CCTCACAACC TGCTGTTTCT TCTTACCATT TCCATCTTCC TGGGGCTGGG
     TCCGCCCGTC GTCGACGTCC GACTGGAACG TCGAACCGCC TTACCTGACC GGAGTGTTGG ACGACAAAGA AGAATGGTAA AGGTAGAAGG ACCCCGACCC
   1                                             M  D  W  P  H  N  L  L  F  L  L  T  I  S  I  F  L  G  L  G

 101 CCAGCCCAGG AGCCCCAAAA GCAAGAGGAA GGGGCAAGGG CGGCCTGGGC CCCTGGCCCC TGGCCCTCAC CAGGTGCCAC TGGACCTGGT GTCACGGATG
     GGTCGGGTCC TCGGGGTTTT CGTTCTCCTT CCCCGTTCCC GCCGGACCCG GGGACCGGGG ACCGGGAGTG GTCCACGGTG ACCTGGACCA CAGTGCCTAC
  21 Q  P  R  S  P  K  S  K  R  K  G  Q  G  R  P  G  P  L  A  P  G  P  H  Q  V  P  L  D  L  V  S  R  M

 201 AAACCGTATG CCCGCATGGA GGAGTATGAG AGGAACATCG AGGAGATGGT GGCCCAGCTG AGGAACAGCT CAGAGCTGGC CCAGAGAAAG TGTGAGGTCA
     TTTGGCATAC GGGCGTACCT CCTCATACTC TCCTTGTAGC TCCTCTACCA CCGGGTCGAC TCCTTGTCGA GTCTCGACCG GGTCTCTTTC ACACTCCAGT
  54 K  P  Y  A  R  M  E  E  Y  E  R  N  I  E  E  M  V  A  Q  L  R  N  S  S  E  L  A  Q  R  K  C  E  V  N

 301 ACTTGCAGCT GTGGATGTCC AACAAGAGGA GCCTGTCTCC CTGGGGCTAC AGCATCAACC ACGACCCCAG CCGTATCCCC GTGGACCTGC CGGAGGCACG
     TGAACGTCGA CACCTACAGG TTGTTCTCCT CGGACAGAGG GACCCCGATG TCGTAGTTGG TGCTGGGGTC GGCATAGGGG CACCTGGACG GCCTCCGTGC
  88  L  Q  L  W  M  S  N  K  R  S  L  S  P  W  G  Y  S  I  N  H  D  P  S  R  I  P  V  D  L  P  E  A  R

 401 GTGCCTGTGT CTGGGCTGTG TGAACCCCTT CACCATGCAG GAGGACCGCA GCATGGTGAG CGTGCCGGTG TTCAGCCAGG TTCCTGTGCG CCGCCGCCTC
     CACGGACACA GACCCGACAC ACTTGGGGAA GTGGTACGTC CTCCTGGCGT CGTACCACTC GCACGGCCAC AAGTCGGTCC AAGGACACGC GGCGGCGGAG
 121 C  L  C  L  G  C  V  N  P  F  T  M  Q  E  D  R  S  M  V  S  V  P  V  F  S  Q  V  P  V  R  R  R  L

 501 TGCCCGCCAC CGCCCCGCAC AGGGCCTTGC CGCCAGCGCG CAGTCATGGA GACCATCGCT GTGGGCTGCA CCTGCATCTT CGGAATCACC TGGCCCAGAA
     ACGGGCGGTG GCGGGGCGTG TCCCGGAACG GCGGTCGCGC GTCAGTACCT CTGGTAGCGA CACCGCAGCT GGACGTAGAA GACTTAGTGG ACCGGGTCTT
 154 C  P  P  P  P  R  T  G  P  C  R  Q  R  A  V  M  E  T  I  A  V  G  C  T  C  I  F  O

 601 GCCAGGCCAG CAGCCCGAGA CCATCCTCCT TGCACCTTTG TGCCAAGAAA GGCCTATGAA AAGTAAACAC TGACTTTTGA AAGCAAG
     CGGTCCGGTC GTCGGGCTCT GGTAGGAGGA ACGTGGAAAC ACGGTTCTTT CCGGATACTT TTCATTTGTG ACTGAAAACT TTCGTTC
```

FIGURE 1

<subunit 1 of 1, 180 aa, 1 stop
<MW: 20437, pI: 9.58, NX(S/T): 1
MDWPHNLLFLLTISIFLGLGQPRSPKSKRKGQGRPGPLAPGPHQVPLDLVSRMKPYARME
EYERNIEEMVAQLRNSSELAQRKCEVNLQLWMSNKRSLSPWGYSINHDPSRIPVDLPEAR
CLCLGCVNPFTMQEDRSMVSVPVFSQVPVRRRLCPPPPRTGPCRQRAVMETIAVGCTCIF

SEQ. ID NO:2

Signal peptide:
1-20 MDWPHNLLFLLTISIFLGLG

N-glycosylation site.
75 NSSE

Homologous region to IL-17
96-180.

FIGURE 2

```
3577142          1 GGCAGCAGGGACCAAGAGAGGCACGCTTGCCCTTTTATGACATCAGAGCT
                   eeeeeeeeeeeeeeeeeeeeeeeeeeeeeeeeeeeeeeeeeeeeeeeee
<consen01>       1 GGCAGCAGGGACCAAGAGAGGCACGCTTGCCCTTTTATGACATCAGAGCT

3577142         51 CCTGGTTCTTGCTCCTTGGGACTCTGGGACTTACACCAGTGGCACCCCTG
                   eeeeeeeeeeeeeeeeeeeeeeeeeeeeeeeeeeeeeeeeeeeeeeeee
<consen01>      51 CCTGGTTCTTGCTCCTTGGGACTCTGGGACTTACACCAGTGGCACCCCTG

3577142        101 GCTCNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN-NNNNNNNNNN
AA476704         1                 AATTCGGTACGAGGCTGGGGTTC-AGGCGGGCAGC
2967181          1                      CGGCGGCATCTGGGGTTCCAGGCGGGCAGC
W74558           1                                           AGGCGGGCAN-
                   eeeeeeeeeeeeeee..... .   .   ......... ..........
<consen01>     101 GCTCNNNNNNNNNNAATTCGGTACGAGGCTGGGGTTC AGGCGGGCAGC

3577142        150 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
AA476704        35 AGCTGCAGGCTGACCTTGCAGCTTGGCGGAATGGACTGGCCTCACAACCT
2967181         31 AGCTGCAGGCTGACCTTGCAGCTTGGCGGAATGGACTGGCCTCACAACCT
W74558          11 AGCTGCAGGCTGACCTTGCAGCTTGGCGGAATGGACTGGCCTCACAACCT
AA033733         1                       CTTGGCGGAATGGACTGGCCTCACAACCT
                   ..................................................
<consen01>     150 AGCTGCAGGCTGACCTTGCAGCTTGGCGGAATGGACTGGCCTCACAACCT

3577142        200 NNNNTTTCTTCTTACCATTTCCATCTTCCTGGGGCTGGGCCAGCCCAG-G
AA476704        85 GCTGTTTCTTCTTACCATTTCCATCTTCCTGGGGCTGGGCCAGCC-AGCG
2967181         81 GCTGTTTCTTCTTACCATTTCCATCTTCCTGGGGCTGGGCCAGCCCAG-G
W74558          61 GCTGTTTCTTCTTACCATTTCCATCTTCCTGGGGCTGGGC-AGCC-AG-G
AA033733        30 GCTGTTTCTTCTTACCATTTCCATCTTCCTGGGGCTGGGC-AGCC-AG-G
                   ....++++++++++++++++++++++++++++++++++++++++++++ +
<consen01>     200 GCTGTTTCTTCTTACCATTTCCATCTTCCTGGGGCTGGGCCAGCCCAG G

3577142        249 AGCCCCCAAAAGCAAGAGGAAGGGGCAAGGGCGGCCTG
AA476704       134 AGCCCC-AAAAGCAAGAGGAAGGGGCAAGGGCGGCCTGGGCCC-TGGTCC
2967181        130 AGCCCC-AAAAGCAAGAGGAAGGGGCAAGGGCGGCCTGGGCCCCTGGCCC
W74558         108 AGCCCC-AAAAGCAAGAGGAAGGGGCAAGGGCGGCCTGGGCCCNTGG--C
AA033733        77 AGCCCC-AAAAGCAAGAGGAAGGGGCAAGGGCGGCCTGGGCCC-TGG--C
                   ++++++ ++++++++++++++++++++++++++++++++++.+++.++
<consen01>     249 AGCCCC AAAAGCAAGAGGAAGGGGCAAGGGCGGCCTGGGCCCCTGGTCC

AA476704       182 CTGGCCCTCACCAGGTGCCACTGGACCTGGTGTCACGGATGAAACCGTAT
2967181        179 CTGGCCCTCACCAGGTGCCACTGGACCTGGTGTCACGGATGAAACCGTAT
W74558         155 CTGGCC-TCACCAGGTGCCACTGGACCTGGTGTCACGGATGAAACCGTAT
AA033733       123 CTNGGCCTCACCAGGTGCCACTGGACCTGGTGTCACGGATGAAACCGTAT
                   ++.+.++++++++++++++++++++++++++++++++++++++++++++
<consen01>     298 CTGGCCCTCACCAGGTGCCACTGGACCTGGTGTCACGGATGAAACCGTAT

AA476704       232 GCCCGCATGGAGGAGTATGAGAGGAACATCGAGGAGATGGTGGCCCAGCT
2967181        229 GCCCGCATGGAGGAGTATGAGAGGAACATCGAGGAGATGGTGGCCCAGCT
```

FIGURE 3A

```
W74558           204 GCCCGCATGGAGGAGTATGAGAGGAACATCGAGGAGATGGTGGCCCAGCT
AA033733         173 GCCCGCATGGAGGAGTATGAGAGGAACATCGAGGAGATGGTGGCCCAGCT
DNA47332.init      1                       GGGAACATCGAGGAGATGTTGGCCCAGCT
W74664.RC          1                       GGGAACATCGAGGAGATGTTGGCCCAGCT
                     ++++++++++++++++++++.+++++++++++++++++ ++++++++++
<consen01>       348 GCCCGCATGGAGGAGTATGAGAGGAACATCGAGGAGATGTTGGCCCAGCT

AA476704         282 GAGGAACAGCTCAGA-GCTGGCCCAGAGAAA-GTGTGAGGTCAA-CTTGC
2967181          279 GAGGAACAGCTCAGA-GCTGGCCCAGAGAAA-GTGTGAGGT
W74558           254 GAGGAACAGCTCANAAGCTGGCCCAGAGAAA-GTGTGAGGTCAA-CTTGC
AA033733         223 GAGGAACAGCTCAGA-GCTGGCCCAGAGAAATGTGTGAGGTCAA-CTTGC
DNA47332.init     30 GAGGAACAGTTCAGA-GCTGGCCCAGAGAAA-GTGTGAGGTCAAACTTGC
W74664.RC         30 GAGGAACAGTTCAGA-GCTGGCCCAGAGAAA-GTGTGAGGTCAAACTTGC
AA443286.RC        1                                    GAGGTCAA-CTTGC
                     +++++++++ +++.+ +++++++++++++++ ++++++++++++ +++++
<consen01>       398 GAGGAACAGTTCAGA GCTGGCCCAGAGAAA GTGTGAGGTCAA CTTGC

AA476704         329 AGCTGTGGATGTCCAA-CAA-GA-GGAGCCTGTCTCCCT-GGGGCTACA-
W74558           302 AGCTGTGGATGTCCAA-CAA-GAAGGAGCCTGTCTCCCTTGGGGCTACAA
AA033733         271 AGCTGTGGATGTCCAA-CAA-GA-GGAGCCTGTCTCCCT-GGGGCTACA-
DNA47332.init     78 AGCTGTGGATGTCCAAACAAAGA-GGAGCCTGTCTCCCT-GGGGCTACA-
W74664.RC         78 AGCTGTGGATGTCCAAACAAAGA-GGAGCCTGTCTCCCT-GGGGTTACA-
AA443286.RC       14 AGCTGTGGATGTCCAA-CAA-GA-GGAGCCTGTCTCCCT-GGGGCTACA-
                     ++++++++++++++++ +++ ++ +++++++++++++++ ++++.++++
<consen01>       445 AGCTGTGGATGTCCAA CAA GA GGAGCCTGTCTCCCT GGGGCTACA

AA476704         374 GCAT
W74558           350 GCATCAACCACCGACCCC-AGCCGTATCCCCGTGGGACC-TTGCCGGGAC
AA033733         316 GCATCAACCAC-GACCCC-AGCCGTATCCCCGTGGACCTGCCGGAGNG--
DNA47332.init    125 GCATCAACCAC-GACCCC-AGCCGTATCCCCGTGGACCT-CCGGAGGCAC
W74664.RC        125 GCATCAACCAC-GACCCCNAGCCGTATCCCCGTGGACCT-NCGGAGGCAC
AA443286.RC       59 GCATCAACCAC-GACCCC-AGCCGTATCCCCGTGGACCT-CCGGAGGCAC
                     ++++++++++ ++++++ ++++++++++++++..+. ..+..+..++
<consen01>       490 GCATCAACCAC GACCCC AGCCGTATCCCCGTGGACCT CCGGAGGCAC

W74558           398
AA033733         362 GTNGCCTGTGTCTGGGCT-GTGTGAACCCCTTCACCATGCAGGAGGACCG
DNA47332.init    172 GGTGCCTGTGTCTGGGCTTGTGTGAACCCCTTCACCATGCAGGAGGACCG
W74664.RC        173 GGTGCCTGTGTCTGGGCTTGTGTGAACCCCTTCACCATGCAGGAGGACCG
AA443286.RC      106 GGTGCCTGTGTCTGGGCT-GTGTGAACCCCTTCACCATGCAGGAGGACCG
                     +..++++++++++++++++++++++++++++++++++++++++++++++++
<consen01>       537 GGTGCCTGTGTCTGGGCTTGTGTGAACCCCTTCACCATGCAGGAGGACCG

AA033733         411 CAGCATGGTGAGCGTGCCGGTGTTCAGCCAGGGTTCCTG
DNA47332.init    222 CAGCATGGTGAGCGTGCCGGTGTTCAGCCAGG-TTCCTGTGCGCCGCCGC
W74664.RC        223 CAGCATGGTGAGCGTGCCGGTGTTCAGCCAGG-TTCCTGTGCGCCGCCGC
AA443286.RC      155 CAGCATGGTGAGCGTGCCGGTGTTCAGCCAGG-TTCCTGTGCGCCGCCGC
                     ++++++++++++++++++++++++++++++++ +++++++++++++++++
<consen01>       587 CAGCATGGTGAGCGTGCCGGTGTTCAGCCAGG TTCCTGTGCGCCGCCGC

DNA47332.init    271 CTCTGCCCGCCACCGCCCCGCACAGGGCCTTGCCGCCAGCGCGCAGTCAT
W74664.RC        272 CTCTGCCCGCCACCGCCCCGCACAGG-CCTTGCC-CCAGCGCGCAGTCAT
AA443286.RC      204 CTCTGCCCGCCACCGCCCCGCACAGGGCCTTGCCGCCAGCGCGCAGTCAT
                     ++++++++++++++++++++++++++++++++++++++++++++++++++++
```

FIGURE 3B

```
<consen01>        636 CTCTGCCCGCCACCGCCCCGCACAGGGCCTTGCCGCCAGCGCGCAGTCAT

DNA47332.init     321 GGAGACCATCGCTGTGGGCTGCACCTGCATCTTCTGAATCGACCTGGCCC
W74664.RC         320 GGAGACCATCGCTGTGGGCTGCACCTGCATCTTCTGAATCGACCTGGCCC
AA443286.RC       254 GGAGACCATCGCTGTGGGCTGCACCTGCATCTTCTGAATC-ACCTGGCCC
                      ++++++++++++++++++++++++++++++++++++++++++++++++++++
<consen01>        686 GGAGACCATCGCTGTGGGCTGCACCTGCATCTTCTGAATCGACCTGGCCC

DNA47332.init     371 AGAAGCCAGGCCAGCAGCCCGAGACCATCCTCCTTGCACCTTTGTGCCAA
W74664.RC         370 AGAAGCCAGGCCAGCAGCCCGAGACCATCCTCCTTGNACCTTTGTGCCAA
AA443286.RC       303 AGAAGCCAGGCCAGCAGCCCGAGACCATCCTCCTTGCACCTTTGTGCCAA
                      ++++++++++++++++++++++++++++++++++.++++++++++++++
<consen01>        736 AGAAGCCAGGCCAGCAGCCCGAGACCATCCTCCTTGCACCTTTGTGCCAA

DNA47332.init     421 GAAAGGCCTATGAAAAGTAAACACTGACTTTTGAAAGCAAAAAAA
W74664.RC         420 GAAAGGCCTATGAAAAGTAAACACTGACTTTTGAAAGCAAG
AA443286.RC       353 GAAAGGCCTATGAAAAGTAAACACTGACTTTTGAAAGCAAAAAAA
                      ++++++++++++++++++++++++++++++++++++++++++.++++
<consen01>        786 GAAAGGCCTATGAAAAGTAAACACTGACTTTTGAAAGCAAAAAAA
```

FIGURE3 C